REPLICATION OF THE LINEAR PLASMIDS OF *KLUYVEROMYCES LACTIS*: ANALYSIS OF THE PUTATIVE DNA POLYMERASES.

CHARLES A. AMBROSE

LEICESTER BIOCENTRE

(Department of Genetics).

Submitted for the degree of
Doctor of Philosophy
of the University of Leicester.

May, 1993

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

CONTENTS.

REPLICATION OF THE LINEAR PLASMIDS OF *KLUYVEROMYCES LACTIS*: ANALYSIS OF THE PUTATIVE DNA POLYMERASES.

CHARLES A. AMBROSE.

ABSTRACT.

The dairy yeast *Kluyveromyces lactis* contains two cytoplasmic linear plasmids pGKL1 (K1) and pGKL2 (K2) which are associated with a killer and immunity phenotype. DNA sequence analyses have revealed that both K1 and K2 contain ORFs (ORFs 1 and 2, respectively) with the potential to encode Class B DNA polymerases (Dpols). This is consistent with a growing body of evidence which indicates that many 'linear plasmids' are dependent for their replication upon self-encoded DNA and RNA polymerases.

Attempts to over-express the native Dpol genes of both K1 and K2, in *E. coli*, were largely unsuccessful. Fragments, ranging from 390 bp to the entire 3 kb gene, were cloned into a variety of expression vectors, but no protein products were observed. The lack of expression arose due to problems at the transcriptional, translational and post-translational level, reflecting the difficulty *E. coli* probably experiences expressing DNA of such a high A+T content. Experiments using gene fusions revealed a dramatic reduction in the level of product yield when the native K-plasmid DNA was coupled to the highly expressed amino terminal stabilising moiety. *In vitro* transcription data also revealed that transcription of the native genes appeared to be prematurely terminating, probably due to the occurrence of fortuitous transcriptional terminator sequences in the A+U rich mRNA.

A 168 bp fragment of the extreme 5' end of the putative K2 Dpol was chemically resynthesised, incorporating an optimal codon bias for high level expression in *E. coli*. The gene fragment was designed such that a second section of the gene could easily be added at a later date, doubling the size of the potential product. This fragment was cloned into a fusion vector which directed the expression of heterologous genes as C-terminal fusions with the 27.5 kDa enzyme glutathione S-transferase (GST). Upon induction, strains bearing this plasmid expressed the GST-Dpol fusion protein to over 10% of total cellular protein. The addition of the extra 54 amino acids to the GST was, however, sufficient to render most of the fusion product insoluble. In the absence of a stabilising conjugant peptide, the small resynthesised gene was transcribed, but a protein product failed to accumulate.

Peptides corresponding to two potentially antigenic sites, within the N-terminus of the K2 Dpol, were chemically synthesised. The two 8-residue peptides were coupled to tetravalent Multiple Antigen Peptide cores, and were used to immunise chickens. The peptides, in this form, however, failed to elicit an immune response from the recipients.

The presence of covalently attached terminal proteins and the cytoplasmic location of these plasmids has severely impaired the ability to manipulate these plasmids. However, homologous recombination *in vivo* has been developed and used to incorporate selectable markers, to disrupt plasmid-borne genes and to re-introduce modified versions of endogenous genes. An affinity tag, consisting of six adjacent histidine residues was incorporated into the 3' end of the putative K1 Dpol gene. The modified gene was then coupled to a selectable marker and successfully re-introduced into K1.

A *K. lactis* nuclear vector was constructed which expressed an antisense RNA complementary to the extreme 5' end of the putative K1 Dpol gene. Despite constitutive high level expression of this RNA, the replication of plasmid K1 appeared to be unaffected.

ABBREVIATIONS.

ATP	adenosine 5' triphosphate;
Ad	adenovirus;
Amp^R	ampicillin resistance;
AS	antisense;
bp	base pairs;
BSA	bovine serum albumin;
DNA	deoxyribonucleic acid;
dNTP	deoxyribonucleotide triphosphate;
DTT	dithiothreitol;
Dpol	DNA polymerase;
DBP	DNA binding protein;
ds	double stranded;
C-terminal	carboxy-terminal;
EDTA	ethylene diamine tetra acetic acid;
GST	glutathione-S-transferase;
ITR	inverted terminal repeat;
ICF	initiation complex formation;
IPTG	isopropyl-ß-D-thiogalactopyranoside;
IgG	immunoglobulin G;
kb	kilobase pairs;
kDa	kilodaltons;
K1	pGKL1;
K2	pGKL2;
K-plasmids	pGKL1 & pGKL2
Km^R	kanamycin resistance;
mRNA	messenger RNA;
MOPS	3[N-morpholino]propane-sulphonic acid;
NF	nuclear factor;

N-terminal	amino terminal;
nt	nucleotide;
ORF	open reading frame;
OD	optical density;
ori	origin of DNA replication;
PAGE	polyacrylamide gel electrophoresis;
PCR	polymerase chain reaction;
pTP	precursor terminal protein;
PEG	polyethylene glycol;
TP	terminal protein;
RNA	ribonucleic acid;
Rpol	RNA polymerase;
ss	single stranded;
SDS	sodium dodecyl sulphate;
SDM	site directed mutagenesis;
SP	synthetic peptide;
Tris	tris (hydroxymethyl) aminomethane;
ts	temperature sensitive;
UCS	upstream conserved sequence;
UV	ultra violet light;
X-GAL	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside;
ø10	bacteriophage T7 gene 10 promoter;
øT	bacteriophage T7 gene 10 terminator;

ACKNOWLEDGEMENTS.

Its been a long and difficult (but generally enjoyable) three years and I'd like to say thanks to everybody who's worked here with me, but specially to: **Pete Meacock** for offering me the opportunity in the first place & for his subsequent support & encouragement; **Pat Wilson**, for media preparation and lots more besides (!); **Dicky Hather**, for help with the sequencing; **Bernie Jordan** & **Rob Mount** for getting me settled in when I first arrived (in the lab & in the Sports Centre); **Keith Lindsey** & **Jen Topping** for the parties, & the Friday nights down at the Marquis; **Pete Brophy, Raffael Schaffrath** & **Surinder Soond** for their plasmids, solutions, encouragement etc. etc.; **Tony Maxwell** (& various members of his lab) for help & critical advice; **Phil Mallinder** (ICI Joint lab) for the advice on gene resynthesis; **Steve Saville** (Genetics) for advice with the chickens/ antibodies; all members, past & present, of the Biocentre 5-aside football team; Status Quo, for all the great music over the years & Ford Escorts, for stopping me getting bored! (but not totally hacked off!); and, of course, **Jane Gilbert**, for her advice, support, patience & friendship.

But above all I want to dedicate this work to two people who have always encouraged & supported me, whatever I was doing. Firstly to my MUM, and secondly to my **Auntie Eileen.**

CHAPTER 1.

INTRODUCTION.

1.1 DNA Replication.

Genome replication is arguably the most fundamental process carried out by living organisms. The survival of an organism and the subsequent propagation of its genetic material to progeny, depends upon the exact duplication of its genome. This is an extremely complex process, requiring the precise interaction of a large number of proteins, and as such has to be very tightly co-ordinated and regulated. The informational content of nucleic acids is determined by the sequential arrangement of the four nucleotide (nt) bases in the chain. There are two such chains in DNA, which are complementary in sequence and are maintained in a helical structure about each other, the so-called 'double helix'. This structure exists due to the specific pairing which occurs between nt in the complementary DNA strands. Each of the four nt types only pairs with one of the other three and this is determined by the precise structural geometry of the appropriate nt as complementary units; the stabilising hydrogen bonds are only established by the exact matching of the two nt.

In general terms, DNA replication can be divided into three stages: initiation, elongation and termination. Initiation is the control point and involves the establishment of the replicating mechanism at a specific DNA sequence in the genome, the origin of replication (*ori*). The proteins involved in replication are assembled here, in a complex (the replisome) which subsequently translocates along the template strand, during the next stage, elongation. This involves extensive polymerisation of the monomer deoxyribonucleotide triphosphates (dNTPs), in a template-directed manner. The final stage of DNA replication is termination, the cessation and disassembly of the replicatory machinery.

Although the precise details of DNA replication in prokaryotes and eukaryotes differ, the basic mechanism is ubiquitous and the general enzymatic activities assembled at the replication fork are outlined in Table 1.1. Due to the structure of DNA, replication is complex and a number of basic problems have to be surmounted. Firstly, the double-stranded (ds) arrangement of DNA means that strand separation is a pre-requisite for replication. This is usually accomplished by sequence-specific DNA binding proteins (DBP) and results in the localised melting of the DNA duplex, at the *ori*. Secondly, the chemical polarity of DNA structure imposes a severe constraint upon the mechanism of its replication.

Table 1.1 The enzymatic activities required at the replication fork during DNA replication (Kornberg & Baker, 1992).

DNA helicase.	Separates the ds helix, by breaking the H-bonds between the base pairs, so each strand can act as a template. As the helical structure is energeticaly stable, helicases use energy liberated by the hydrolysis of NTPs.
Dpol.	The main agents of DNA replication. They function in the fast & accurate synthesis of DNA. This is responsible for their complex structure. Multiple species of Dpol have been identified & purified in both prokaryotes & eukaryotes, where specific functions are performed predominantly by one species.
3'-5' exonuclease.	Functions in maintaining the essential high fidelity of replication, by proofreading the newly syntheised DNA. It removes non-hydrogen bonded bases (ie. mis-inserted) from the 3' end of ss DNA chains.
RNase H.	Involved in the hydrolytic removal of RNA primers.
5'-3' exonuclease.	Involved in the hydrolytic removal of RNA primers & post-synthetic DNA repair.
DNA ligase.	Forms phosphodiester bonds between adjacent sugar moieties in the ds helix, sealing the nicks in the DNA backbone. Activity requires a free 3' OH & 5' phosphate termini, ATP for energy & ds DNA in a helical structure.
DNA topoisomerases	Relieve the torsional stress accumulated during DNA replication. There are two classes, type 1 & type 2 (depending upon whether they make ss or ds nicks in the DNA).

The polarity of each strand is defined by the asymmetry of the sugar-phosphate backbone. The basic structure of each DNA strand in the double helix is identical and the two strands of the double helix have the same polarity. The complementary nature of the structure, therefore results in an anti-parallel arrangement; one strand running in a 5'-3' direction, the other in a 3'-5' direction. Simultaneous replication of both template strands in opposite directions, is complicated by the fact that all known DNA polymerases (Dpol) require a 'priming' 3'OH group from which to polymerise dNTPs in a 5'-3' direction. The priming 3'OH group is usually provided by an oligoribonucleotide synthesised by a 'primase' subunit of the Dpol holoenzyme. This problem has been overcome by the evolution of an asymmetrical, semi-discontinuous mechanism. Replication of one (the leading) strand is continuous requiring only a single priming event, but is discontinuous on the other (lagging) strand, requiring multiple priming events and resulting in the formation of short Okazaki fragments. The simultaneous replication of both strands in co-ordinated fashion is now thought to be accomplished by the lagging strand bending back on itself to form a loop, effectively allowing the Dpol to move along and polymerise dNTPs in the same 5'-3' direction (Wang, 1991; Kornberg & Baker, 1992). Furthermore, the Dpol at the replication fork is also thought to exist as an asymmetric dimer, one molecule replicating each strand. This can either be in the form of a heterodimer where different core molecules replicate each strand (as with eukaryotic Dpol-α and -δ), or as homodimers where the accessory factors complexed with the core molecule define which strand each Dpol replicates (as in *Escherichia coli* Dpol III). The third major obstacle, is that genomes are often organised in to complex higher order domains which have to be undone as replication proceeds.

Prokaryotic and eukaryotic DNA viruses have been very useful models for understanding DNA replication. Many have evolved mechanisms to escape the replicatory control imposed by the host, either by using different mechanisms or by synthesising specific viral proteins that direct viral DNA replication. Other viruses do, however, utilise cellular mechanisms to propagate their genome. Many proteins which replicate viral DNA, therefore, probably play a similar role in host genome replication. A basic understanding of DNA replication in these systems has yielded much information concerning this process in higher organisms. The enzymology of DNA replication was first elucidated in the prokaryote *E. coli* by the study of

several of its bacteriophages, most notably the single stranded (ss) phages M13 and øX174, and the ds T4 and T7 phages. The details of eukaryotic DNA replication were subsequently studied using Herpes Simplex (HSV), Simian Virus type 40 (SV40) and Adenovirus (Ad). Both the prokaryotic phages M13 and øX174 use the host replicative machinery and encode only a few of their own replicatory requirements, as does the eukaryotic SV40 virus. In contrast, T4 and T7 DNA replication is catalysed by a complex of essential viral genes products, as is the case with the eukaryotic HSV. The characterisation of the proteins involved in viral and host DNA replication has allowed the development of cell-free DNA replication systems. The first such *in vitro* DNA replication system was developed using purified proteins from *E. coli* and the phages T4 and T7. However, a cell free replication system for mammalian cells was only very recently developed.

Prokaryotes are now known to contain three different Dpols. Dpol I of *E. coli* was the first to be discovered and has since been the most extensively studied. Two further Dpols were subsequently discovered in Dpol I mutants (reviewed by Balbinder & Walden, 1991; Kornberg & Baker, 1992). Dpol III is now known to be the main replicative enzyme. The holoenzyme consists of at least 22 subunits with a total molecular weight approaching 900 kDa. The first eukaryotic Dpol was discovered in calf thymus, since which five different Dpols (α, β, δ, γ and ε) have been identified from a range of eukaryotic organisms, the structure of which are similar in higher and lower eukaryotes (reviewed by Wang, 1991; Kornberg & Baker, 1992). Dpol-α and Dpol-δ are now thought to be the main replicative enzymes.

1.2 The Fidelity of DNA Replication.

The maintenance of low mutation rates, in prokaryotic replication systems, is achieved by the sequential operation of three fidelity mechanisms (Eckert & Kunkel, 1991; Echols & Goodman, 1991; Balbinder & Walden, 1991): selection of the correct dNTP during polymerisation; 3'-5' exonucleolytic removal of mis-incorporated bases from the end of the growing chain prior to further elongation (proof-reading); excision of mis-matched dNTPs after passage of the replication fork (mis-match repair). In eukaryotes only base selection and 3'-5' exonucleolytic mechanisms are currently known to operate.

Base selection is achieved by the precise structural complementarity of nt pairing in the double helix; all other combinations have a different and hence less favourable geometry. The active site of the Dpol, therefore, exerts a selectivity for the optimal nt arrangement and this results in the correct dNTP being inserted much more efficiently than the incorrect dNTP. The mechanisms for such base discrimination are not clear at present. Insertion of an inappropriate nt may, however, be manifest in a more rapid dissociation of the incorrect dNTP from the Dpol active site or slower subsequent phosphodiester bond formation. The normal hydrogen-bonding between nt can be disrupted by spontaneous fluctuations in the electronic structure (tautomerism) of the bases and this can lead to base mis-pairing, and hence constitutes a cause of spontaneous mutations. Mis-pairing errors have been estimated to occur with a frequency of 10^{-1} to 10^{-2} mis-incorporations per bp replicated; but data generated *in vivo* suggests a frequency of 10^{-7} to 10^{-11}. The basis for this disparity in frequencies is the operation of specific error correcting components in the replicatory apparatus.

The 3'-5' exonucleolytic activity of Dpols was originally thought to be exclusive to prokaryotes, but is now known to be widespread in prokaryotic, eukaryotic and viral Dpols. This activity has, however, only really been characterised in *E. coli* Dpol I, so its contribution to the fidelity of eukaryotic DNA replication remains largely unknown. Studies with *E. coli* Dpol I have demonstrated that the active site lies in the N-terminal part of the Klenow fragment (the large C-terminal proteolytic product which also contains the polymerase activity), in a structurally distinct domain that has binding sites for divalent metal ions, ss DNA and dNTPs. Evidence generated with *E. coli* Dpol I and III, and T4 Dpol (Reddy *et al.*, 1992) suggest that proof-reading probably operates on the basis of melting capacity; a mis-paired base at the 3' end of the elongating DNA strand more often being in a ss configuration. This, coupled with the relatively slow addition of the subsequent nt (compared to a previously correctly inserted base) renders the defective template more susceptible to the exonuclease. The importance of the editing mechanism to the fidelity of replication has been defined both *in vivo* and *in vitro*. T4 Dpol mutants with a reduced exonuclease activity, correlated with higher mutation rates. Furthermore, comparison of 3'-5' exonuclease-containing and exonuclease-deficient Klenow fragments demonstrated a 7-

fold difference in the generation of base substitution mutations. The thermostable *Vent* polymerase (from *Thermococcus litoralis*) has an intrinsic 3'-5' exonuclease resulting in an estimated base substitution error rate of 1 in 31,000 at 1 mM dNTP, similar to that observed with Klenow, but 5-fold lower than that generated by the exonuclease-deficient *Taq* polymerase (from *Thermus aquaticus*) under similar conditions (Eckert & Kunkel, 1991). However, the proof-reading system has a substantial cost to the cell; it has been estimated that, in *E. coli* Dpol III and T4 Dpol, 5-15% of correctly inserted bases are removed. This therefore suggests that the proof-reading and polymerase activities are in direct competition, and has recently been demonstrated *in vitro* for the ø29 Dpol (Garmendia *et al.*, 1992). A dNTP concentration of 20-30 nM was normally required to allow predominance of the polymerase activity, but selective inactivation by site directed mutagenesis (SDM) of the proof-reading activity resulted in a 350-fold reduction in the dNTP concentration required for elongation.

1.3 Structural and Functional Conservation Among DNA Polymerases.

Biochemical and genetic analysis of viral DNA replication has revealed that certain mutations in the Dpol genes could confer resistance to certain antiviral drugs, such as dNTP analogs (eg. acyclovir and aphidicolin) and pyrophosphate analogs (eg phosphonoacetic acid). As these drugs mimic the natural substrates, mutations that alter the interaction of the enzyme with these analogs may define functionally important domains or residues. Mapping and characterisation of HSV and ø29 Dpol mutants has revealed that many of these lesions were concentrated in the C-terminal portion of the protein, often in amino acid sequences which exhibited a remarkable degree of similarity between different Dpols (Gibbs *et al.*, 1985; Matsumoto *et al.*, 1986; Larder *et al.*, 1987; Hall, 1988; Marcy *et al.*, 1990). Comparison with a wider range of Dpols subsequently revealed that these sequences were also conserved in apparently unrelated enzymes (eg. Vaccinia, Ad and ø29). These domains were designated I-VI (or VII; Morrison *et al.*, 1991) according to their degree of conservation, region I being the most highly conserved. The significance of these similarities was underlined by their similar linear and spatial organisation; the motifs always occurred in the same linear order (VII-IV-II-VI-III-I-V) and occupied similar spatial co-ordinates in the primary sequence. The relative positions of the three most highly conserved 'polymerase' domains, in selected enzymes, are shown in Table 1.2. Regions I-IV have been implicated in the polymerase function, regions

Table 1.2 Location of the Three Conserved 'Polymerase' Domains in the Primary Sequence of Selected Dpols (adapted from Bernad *et al.*, 1987). Regions I-III refer to the C-terminal sequence similarities, region I being the most highly conserved. The % figures indicate the position of each region along the peptide, reletive to the N-terminus of the protein. Note that the spacing of the three regions is remarkably similar in each proteins and may reflect three dimensional structural conservation.

Host.	enzyme size (amino acids)	Relative position (%) of the conserved domains along the primary sequence.		
		Region II (consensus SLYP)	Region III (consensus NS-YG-F)	Region I (consensus YGDTD)
S1	917	53	68	78
K1	995	66	80	88
K2	998	60	75	90
Ad	1056	52	67	82
ø29	575	44	68	79
PRD1	553	41	63	78
T4	896	46	63	69
Vaccinia	938	56	69	78
HSV-2	1240	59	66	72
VZV	1194	58	66	71
EBV	1015	58	68	74
HCMV	1242	58	66	73
M2	572	39	63	78
huDp-α	1462	57	64	68
S.cerevisiae DpI	1468	58	66	68
pCIK1	1097	55	77	82
S.solfactaricus	882	56	67	73
pSKL	999	64	78	88
Kalilo	970	64	79	88
pAL2-1	1198	56	66	76

II, III and V in dNTP binding (Wang *et al.*, 1989). Not all Dpols, however, contain all seven sequences; the Ad Dpol only contains five, the ø29 Dpol only four. The majority of HSV Dpol mutations conferring drug resistance mapped to single nt changes in conserved regions II and III, and resulted in conservative amino acid substitutions (Gibbs *et al.*, 1985; Knopf, 1987). Similar results have been obtained by analysis of phosphonoacetic acid and aphidicolin resistant mutants in Vaccinia. Some mutations conferring resistance to the same drug mapped to conservative amino acid substitutions at least 228 residues apart, and were thought to define sequences which interacted to form a common functional domain (Gibbs *et al.*, 1985). Similarly, the characterisation of four independently isolated ø29 temperature sensitive (*ts*) Dpol (*ts*2) mutants revealed that although three were due to same site mutations (Blasco *et al.*, 1990), the fourth site mapped 138 amino acids away. Mutations which conferred resistance to one drug also often conferred simultaneous resistance to other Dpol inhibitors, leading to speculation that these sites defined overlapping domains or active sites involved in binding of these drugs, and hence, the natural substrates (Larder *et al.*, 1987; Matsumoto *et al.*, 1990).

Mutations within the most highly conserved region, I (consensus sequence YGDTDS), in the HSV (Marcy *et al.*, 1990) or ø29 Dpol (Bernad *et al.*, 1990a & b) were either found to be lethal, to severely impair viral growth, or to confer altered drug sensitivity. On the basis of secondary structure predictions (Argos, 1988) and comparisons with a similar motif in *E. coli* Dpol I (Delarue *et al.*, 1990) and other metal and dNTP binding sites, region I is thought to be involved in Mg^{2+} binding, in the dNTP interacting domain of the polymerase activity. A motif similar to region I is also highly conserved among RNA-directed RNA and Dpols (Kamer & Argos, 1984) and mutagenesis data has confirmed the critical nature of this sequence in these enzymes. Five of six mutants generated in this sequence in the HIV-1 reverse transcriptase were inactive *in vitro*; the single active enzyme exhibited a 2-3 fold increased resistance against both NTP and pyrophosphate analogs. Some evidence, generated by SDM, also suggests the pyrophosphorolytic activity of Dpol (the reversal of polymerisation) is involved with region I (Blasco *et al.*, 1992). The functional importance of this YGDTDS sequence in the initiating and extending activity of the enzyme has been demonstrated by SDM, in both the Ad and ø29 Dpols (Bernad *et al.*, 1990 a & b; Joung *et al.*, 1991). Both

conservative and non-conservative amino acid changes resulted in reduced enzyme activity; alterations at certain sites completely abolishing both initiating and elongating activities. These effects were also specific for the polymerase activity; in each case the other activities of the enzyme and its ability to interact with other replicatory proteins was unaffected. Furthermore, all these results suggest that there is no physical separation of the initiation and elongation activities.

The functional and structural conservation among Dpols is now known to extend beyond the 'polymerase' domains. Crystallographic and mutational evidence has suggested that the 3'-5' exonuclease active site of *E. coli* Dpol I is composed of a group of acidic amino acid residues clustered around two divalent metal ions. Computer analyses of other Dpol sequences, using the *E. coli* Dpol I proof-reading domain as a model, revealed three further conserved sequences within the N-termini of many Dpols, again in the same relative linear order (Bernad *et al.*, 1987). These were termed *Exo*I (the furthest upstream), II and III. *Exo*I and II lie within conserved 'polymerase' region IV; *Exo*III lies 50-150 residues downstream of these. On the basis of these sequence resemblances, various functions have been assigned to these motifs, including metal, ss DNA, dNTP and substrate binding and direct interaction with the bond to be hydrolysed. The enzymatic importance of these motifs has been demonstrated by SDM of the ø29 Dpol (Bernad *et al.*, 1989 & 1990a & b; Soengas *et al.*, 1992). Alterations in either the *Exo*I or *Exo*II domains selectively abolished 3'-5' exonuclease activity, but all other enzymic activities of the protein (eg. ss DNA binding, processivity or protein-primed initiation) were unaffected. As expected, these mutations produced enzymes which exhibited an increased frequency of misinsertion and extension of Dpol errors. Similarly, substitution mutations in the corresponding regions of the *Saccharomyces cerevisiae* Dpol II (Morrison *et al.*, 1991) also drastically and selectively reduced exonuclease activity. Yeast strains expressing this exonuclease-deficient enzyme exhibited a 22-fold increase in the spontaneous mutation rate. The T5 (Leavitt & Ito, 1989) and *Plasmodium falciparum* (Fox & Bzik, 1991) Dpols also share both proof-reading and polymerase sequence similarities with *E. coli* Dpol I, as does the thermostable Dpol from the thermophilic archaebacterium *Sulfolobus solfactaricus* (Pisani *et al.*, 1992). *Taq* Dpol contains only the C-terminal polymerase motifs, the absence of the N-terminal proof-reading domains

reflecting the lack of a proof-reading capability. The N-terminal portion of this enzyme does however, resembles the 5'-3' exonuclease domain of *E. coli* Dpol I (Longley *et al.*, 1990).

The apparent conservation of functional regions within a wide range of Dpols has led to much speculation concerning the evolution of these enzymes. Those bearing the conserved motifs presumably being of common origin, having diverged from a primordial progenitor. Those Dpols lacking such conserved sequences may, however, reflect convergent evolution, as a result for the quest for a common function. Furthermore those enzymes with a modular structure may have arisen by fusion of ancestral genes with closely related functions (eg. the proof-reading, polymerase and 5'-3' exonuclease domains of *E. coli* Dpol I being segregated into discrete sub-domains). Some experimental evidence, generated with ø29 T4 and *E. coli* Dpols, supports a modular structure for the Dpols containing these sequence motifs. For example, monoclonal antibodies with the ability to specifically immunoneutralise each of the three enzymic activities of *E. coli* Dpol I, have recently been generated (Ruscitti *et al.*, 1992). There is, however, also evidence to the contrary. Some mutations in the *exo*III motif of the ø29 Dpol have been found to reduce the replicative ability of the Dpol due to a defect in the strand displacement activity. This suggests that this activity also resides in the N-terminus of the enzyme and overlaps the proof-reading active site (Soengas *et al.*, 1992). The generation of conservative amino acid substitutions in the predicted proof-reading active site of HSV Dpol were also found to impair the ability of the virus to replicate its DNA, but this was due, unexpectedly, to a deficiency in polymerase activity. Furthermore, the functional analysis of linker insertion mutations in the Ad Dpol demonstrated that essential regions mapped across the whole protein and were not limited to the regions of homology shared with other Dpols (Chen & Horowitz, 1989). There may, therefore, be some overlap of enzymic active sites, the conserved sequences reflecting important roles played by these regions in several activities (eg. template, metal or dNTP binding).

Such an observation, a mutation in one of the putative N-terminal proof-reading motifs of the T4 Dpol having little effect upon proofreading activity, prompted the significance of these conserved sequences to be called into question (Reha-Krantz, 1992). This view was, however, strongly contested by a re-alignment of these conserved sequences in thirty three

DNA-dependent Dpols (Blanco *et al.*, 1992a). The aforementioned T4 Dpol mutation was subsequently revealed to map just outside the *Exo*I motif, in a region predicted not to be critical for proof-reading activity.

The accumulated data therefore suggests that these sequence similarities represent important functional regions within the protein. Furthermore, the fundamental nature and ubiquity of genome replication (ie. the functional conservation) is thought to be reflected by a widespread structural conservation at the enzymological level. This has been supported by analyses of an increasing number of Dpol gene sequences.

Although Dpols are specifically classed within prokaryotic and eukaryotic kingdoms (eg. as α–ε and I-III, respectively), some of these enzymes are also occasionally referred to as α-like or class A or B, depending on their resemblance to *E. coli* Dpol I or eukaryotic Dpol-α, (Jung *et al.*, 1987a & b). This nomenclature, together with the rapidly accumulating number of Dpol sequences available, has resulted in the classification of these enzymes becoming increasingly ambiguous. Following the sequencing and characterisation of *E. coli* Dpol II (Chen *et al.*, 1990) a broad classification, based on evolutionary relationships (to *E. coli* Dpol I, II and III) which are reflected in amino acid sequence resemblances has recently been proposed (Ito & Braithwaite, 1991). Therefore, what were formerly called α-like are now called Class B (Table 1.3). It is interesting to note that nearly all the Dpols thought to be involved in a protein primed replication mechanism are of the Class B type.

1.4 Replication of Linear DNA Molecules.

All known Dpols have an absolute primer requirement, a 3' OH group upon which to attach subsequent dNTPs. In the majority of cases this is provided by RNA primers or ss nicks in ds DNA. As genome replication must result in the synthesis of an exact duplicate molecule, the priming of extreme terminal regions of linear DNA molecules is, therefore, problematic. A primer usually consists of a short (6-10 nt) oligoribonucleotide complementary to the ss template from which Dpol can extend. Assuming that a single terminal ribonucleotide is sufficient for priming, completion of replication and hydrolytic removal of the primer would result in a daughter strand one nt shorter than the parental strand. Therefore, unless a mechanism for completing the 5' end of newly synthesised DNA strands or a Dpol which

Table 1.3 Class B DNA polymerases, as defined by their sequence & biochemical similarities to *E. coli* Dpol II (adapted from Ito & Braithwaite, 1991). * denotes protein primed systems. (@) cited in Ito & Braithwaite, 1991

Bacterial.	**size (amino acids).**	**MW (kDa).**	**Reference.**
E.coli Dpol II			Chen *et al.*, 1990
Pyrococcus furiosus			Uemeri *et al.*, 1993
S. solfactaricus	882	100	Pisani *et al.*, 1992
Bacteriophage.			
*ø29	575	66.5	Watabe & Ito, 1983
*PRD1	553	63.3	Jung *et al.*,1987b; Savilahti *et al.*, 1987
T4	896	103.5	Spicer *et al.*, 1988
*M2	572		Matsumoto *et al.*, 1989
Eukaryotic.			
P. falciparum:	1094	121	Fox & Bzik, 1991
S.cerevisiae Dpol I	1468	166.8	Pizzagali *et al.*, 1988
human Dpol-α	1462	165	Wong *et al.*, 1988
S.cerevisiae Dpol II	1462	155	Morrison *et al.*, 1991
S.cerevisiae Dpol III			(@)
Viral.			
Fowlpox virus			(@)
HSV-1	1235	136.52	Gibbs *et al.*, 1985
HSV-2	1240	140	
Baculovirus	984	109	Tomalski *et al.*, 1988
Vaccinia	1006	110	(@)
Epstein-Barr Virus	1015	113	(@)
Varicella-Zoster Virus	1194		(@)
*Ad2	1056	140	(@)
*Ad7			(@)
*Ad12			Shu *et al.*, 1987
HCMV	1242		(@)
Extrachromosomal.			
*S1	917		Paillard *et al.*, 1985
**Kalilo & Maranhar*	970	120	Chan *et al.*, 1991; Court *et al.*, 1991; Myers *et al.*, 1989
*pCIK1	1097		Duvell *et al.*, 1988; Oeser & Tudzinski, 1989
*pAL2-1	1198	140	Hermanns & Osiewacz, 1992
*pAI2	1202		Kistler & Leong, 1986; Kempken *et al.*, 1989
*pMC3-2	902		Rohe *et al.*, 1991
*pEM			Robinson *et al.*, 1991
*pSKL	999		Hishinuma & Hirai, 1991
*K1	995	116.2	Fukuhara, 1987; Jung *et al.*, 1987a
*K2	998	117.4	Tommasino *et al.*, 1988

can initiate synthesis *de novo*, exists, subsequent rounds of replication would yield progressively shorter progeny molecules. This problem has been circumvented in Poxviruses (eg. Vaccinia) and Parvoviruses (eg. SV40) by the presence of terminal palindromic DNA sequences. These lead to the formation of hairpin loops, which prime the replication of terminal sequences. The linear chromosomes of eukaryotes have specialised terminal structures, telomeres (recently reviewed by Blackburn, 1991). Telomeric DNA is synthesised by the ribonucleoprotein telomerase, and generally consist of tandemly repeated DNA sequences 100-1000 bp long. Telomerase is an unusual Dpol in that it has an integral RNA sequence which is used as a template for DNA synthesis (ie. a reverse transcriptase-like activity). Telomeres not only provide priming sites, but also protect critical internal DNA sequences which must be conserved during replication. In the absence of telomeres, linear DNA molecules are prone to degradation by cellular exonucleases, and unprotected termini (generated by restriction enzymes, mechanical breakage or ionising radiation) are highly recombinogenic, and therefore unstable. This phenomenon has been observed *in vivo* in *Drosophila* chromosomes with broken ends, lacking conventional telomeric sequences, and in *Tetrahymena*, *S. cerevisiae* and non-immortalised human cells lacking telomerase. This resulted in a progressive shortening of chromosomes, increased rates of chromosome loss and eventual cell death.

Hairpin or telomeric structures do not exist in the linear DNA molecules of Ad, ø29 or the increasingly recognised number of 'linear-plasmids'. The presence of genomic inverted terminal repeat (ITR) sequences, protein molecules covalently attached to the DNA termini, and self-encoded Dpols, suggested the existence of another replicatory mechanism, where the Dpol is primed by a 3' OH group donated by the protein molecule. Although some viruses (eg. the Parvoviruses SV40) also have covalently attached terminal protein (TP) molecules which are thought to be involved in DNA replication, these do not prime synthesis.

1.5 An Introduction to Linear Plasmids.

When discussing 'plasmids' (autonomously replicating extrachromosomal genetic elements) one, generally, envisages the archetypal small covalently closed circular molecule widely

used in *E. coli* and *S. cerevisiae* molecular genetics. It is now, however, becoming increasingly clear that many organisms also contain 'linear plasmids'.

Extrachromosomal linear ds DNA replicons have been found in organisms as diverse as plants, fungi, bacteria and bacterial and animal viruses; they appear to be particularly widespread in fungi (Table 1.4 & reviews by Meinhardt & Esser, 1987; Salas, 1988, Volkert *et al.*, 1989; Meinhardt *et al.*, 1990; Sakaguchi, 1990; Salas, 1991). These episomes sometimes confer recognised phenotypes upon the host organism including cytoplasmic male sterility in plants, toxin production and immunity in yeast, antibiotic synthesis and antigenic variability in bacteria, and senescence, host specificity and virulence in fungi. Many, however, appear to be cryptic.

These linear DNA molecules range in length from the 1.9 kb in *Fusarium oxysporum* (Kistler & Leong, 1986), to 520 kb in *Streptomyces lasaliensis* (Kinashi & Shinagi, 1987; Kinashi *et al.*, 1987). The majority have two common traits, both of which are thought to constitute essential components of a highly specific protein-primed DNA replication mechanism. These plasmids have protein molecules covalently attached to their 5' DNA termini, ranging in size from 23 kDa of the *S. pneumoniae* phages HB3, HB623 and HB746 (Romero *et al.*, 1990) to 120 kDa in *Kalilo* (Vierula *et al.*, 1990). Their genomes also contain ITRs ranging in length from 6 bp in the bacteriophage ø29 (Salas *et al.*, 1978) to 1,211 bp in pFSC1 of *Nectria haematococca* (Giasson & Lalonde, 1987). Evidence is also accumulating which suggests that many of these elements are dependent for their replication and gene expression upon self-encoded DNA and RNA polymerases (see Tables 1.3 & 1.4). Nevertheless, not all 'linear plasmids', possess ITRs and covalently attached terminal proteins, and those that do not are thought to replicate by mechanisms other than by protein priming.

Although the source of 'linear plasmids' remains obscure, given the many similarities to linear ds DNA viruses of animals and bacteria (for example Ad and ø29) the general consensus of opinion is that these endogenous linear replicons probably share a viral origin.

Table 1.4 An Introduction to Linear Plasmids.

Plasmid.	Host.	Plasmid size (kb).	Comments.	References
Fungi.				
pGKL1 & pGKL2	*Kluyveromyces lactis*	8.9 & 13.5	Cytoplasmic, high A+T content (73%), each has different TPs, K1 encodes toxin & immunity, K2 encodes Rpol, both encode different Dpols	See section 1.7.1
pSKL	*Saccharomyces kluyveri*	14.2	Cytoplasmic, very similar to K2 (overall DNA homology 55%), 71.7% A+T. ITR 483 bp, contains 30 bp sequence repeated 8 times; 15/16 terminal nucleotides identical to K2. Encodes putative R & Dpol. Dpol 62% homologous to K2 Dpol. Appears to be cryptic	Kitada & Hishinuma, 1987; Hishinuma & Hirai, 1991
	Hanselula mrakí	50	Mitochondrial. Cryptic	Wesolowski & Fukuhara, 1981
pPIN1, 2 & 3	*Pichia inositovora*	18, 13 & 10	Cytoplasmic; high A+T content; have TPs; encode toxin, 3rd plasmid not involved toxin/immunity. Very sensitive to its own toxin when plasmids cured	Ligon *et al.*, 1989; Hayman & Bolen, 1991
pPAC-1 & pPAC-2	*Pichia acacia*	13.6 & 7.3	1 encodes immunity; 2 encodes toxin. Both have TPs	Worsham & Bolen, 1990
pSCrl-1, 2 & 3	*Saccharomycopsis crataegensis*	15, 7 & 5	Cytoplasmic; A+T rich; 1 & 2 related to each other *via* sequence, but not to 1; have TPs. Cryptic	Shepherd *et al.*, 1987
pEM, 50H & pMPJ	*Agaricus bisporus*		Mitochondrial, high A+T content (73%), pEM encodes putative D & Rpol. 50H & pEM have 80% DNA homology. Some homology to mt DNA but none to nuclear	Robinson *et al.*, 1991
pAal-1, 2 & 3	*Alternaria alternata*	7, 6.8 & 5	Cytoplasmic, have TPs	Shepherd, 1992
Kalilo	*Neurospora intermedia*	8.6	Mitochondrial. Causes senescence by integration into large (25s) s-u rRNA gene. Encodes Dpol & Rpol. TP 120 kDa. High A+T content (70%), ITR 1361 bp. No homology mt or nuclear DNA	Bertrand *et al.*, 1985 & 1986; Bertrand & Griffiths, 1989; Myers *et al.*, 1989; Vierula *et al*, 1990
maranhar	*Neurospora crassa*	7	Mitochondrial, TPs & ITRs, causes integrative senescence, encodes putative Dpol & Rpol, no significant homology to kalilo, mt or nuclear DNA	Court *et al.*, 1991
pFSC1 & 2	*Fusarium solani*	9.2 & 8.3	Mitochondrial, involved in host pathogenicity & specificity. No significant homology with mt or nuclear DNA.	Samac & Leong, 1988 & 1989
pCF637 & pFQ501	*Ceratocystis fimbriata*	8.2	Have TPs	Giasson & Lalonde 1987
Et7.5L, 2.1L & 2L	*Epichloe typhina*	7.5, 2.1 & 2	Mitochondrial; have TPs; no homology to each other or mt/nuclear DNA. Et2.0L contains ORF with amino acid sequence similarities to some reverse transcriptases, Rpols & Dpols	Mogen *et al.*, 1991
pFOXC1 & 2	*Fusarium oxysporum*	1.9	Mitochondrial, have TPs, no homology with mt or nuclear DNA, involved host specificity & pathogenicity	Kistler & Leong, 1986
pAI1, 2 & 3	*Ascolobus immersus*	7.9, 5.6 & ?	Mitochondrial, have TPs, high A+T content (70%), pAI2 ITR 1200 bp, pAI1 700 bp ITR. *ori* within 1st 500 bp pAI2. pAI2 encodes putative Dpol. No homology to mt DNA. Codon bias typical fungal mt. Cryptic	Francou, 1981; Kempken *et al.*, 1989; Meinhardt *et al.*, 1986
	Physarum polycephalum	13.3 - 18.3	Mitochondrial; no TP or ITR but telomeric seq'uences of variable length	Tokano *et al*, 1991
pAL2-1	*Podospora anserina*		encodes putative Rp	Hermanns & Osiewacz, 1992

pAaL	*Ascosphaera apis*	12	mitochondrial; TPs & ITRs; A+T-rich.	Qin *et al.*, 1993
pLLe1	*Lentinus edodes*	11	Mitochondrial, has telomeres (no TPs), no homology mt/nuclear DNA	Katayose *et al.*, 1990 & 1992
pRS64-1, 2 & 3	*Rhizoctonia solani*	2.7	Hairpin loop termini (no TPs), involved in host pathogenicity & specificity	Hashiba *et al.*, 1984; Miyashita *et al.*, 1990
pMC3-1 & 2	*Morchella conica*	? & 6.044	Mitochondrial, pMC3-2 encodes putative Dpol; ; high A+T content (66%), ITRs 710-713 bp	Meinhardt & Esser, 1984; Meinhardt & Esser, 1987; Rohe *et al.*, 1991
pMH1	*Morchella hortensis*	8.8		
pMS1 & 2	*Morchella esculenta*	4 & 7.5		
pME141-1 & pME141.2	*Morchella elata*	6.7 & 6	pME141.2 identical to pMC3-2	
pCIK1	*Claviceps purpurea*	6.7	Mitochondrial; has TPs; ITRs 327 bp; high A+T (67%), encodes putative R & Dpol, maybe involved in host virulence. Cryptic	Duvell *et al.*, 1988; Oeser & Tudzynski, 1989.
	Tilletia caries, T.laevis, T.contravesa	7.2-7.6	Mitochondrial, all very homologous to each other, but not to mt DNA;	Kim *et al*, 1990
pLPO-1 & 2	*Pleurotus ostreatus*	10 & 9.4	Mitochondrial. Have different TPs, show no homology to each other, mt or nuclear DNA. Appears to be cryptic	Yui *et al.*, 1988, Katayose *et al.*, 1990
E1 & E2	*Gaeumannomyces graminis*	8.4 & 7.2	Mitochondrial, high A+T content	Honeyman & Curvier, 1986
pFSC1 & 2	*Nectria haematococca*	9.2 & 8.3	Mitochondrial, 80 kDa TPs; ITRs of 1211 & 1027 bp. No seq homolog each other.	Samac & Leong, 1989
Plants.				
R1 & 2	*Zea mays*	7.4 & 5.4	Highly homologous to S1 & S2; ITR 200 bp, involved CMS	Levings & Sederoff, 1983
	Zea mays	2.3	Ubiquitous to all maize; mitochondrial, ITRs 170 bp made up short repeats, TPs, plasmids contain normal mt genes, reintegrated plasmid gives fertile pollen; contains ORF with potential to encode 33 kDa product.	Garcia *et al.*, 1988; Leon *et al.*, 1989
S1 & S2	*Zea mays*	6.399 & 5.453	Mitochondrial, have same TPs, share 1.4 kb homology, ITRs 208 bp; S1 encodes Dpol & S2 Rpol. Causes CMS when integrated into mt DNA, S1 & 2 equimolar but copy number 5x greater than mt DNA. The first linear plasmids discovered; extensive homology mt DNA	Pring *et al.*, 1977; Kemble & Thompson, 1982; Schardlt *et al.*, 1985; Traynor & Levings, 1986; Kuzmin & Levchenko, 1987; Kuzmin *et al.*, 1988
	Brassica napus & B.campestris	11.3	mitochondrial; ITR 325 bp; TPs; maybe involved with CMS. No signifc homology mt or nuclear DNA	Palmer *et al.*, 1983; Erickson *et al.*, 1985 & 1989; Turpen *et al.*, 1987.
	Beta maritime	10.4	mitochondrial. Appears to be cryptic	Saumitou-Laprade *et al.*, 1989
N1 & N2	Sorgum	5.7 & 5.3	Mitochondrial; extensive homology to each other & some homolg S1 & S2	Pring *et al.*, 1982
Bacteria.				
pSLA1 & 2	*Streptomyces rochei*	17	614 bp ITR; has TPs; encodes antibiotics,lankacidins. v.similar each other in terms of seq.	Hirochika & Sakaguchi, 1982; Hirochika *et al.*, 1984
pKSL	*Streptomyces lasaliensis*	520	encodes antibiotics lasalocid A & echinomycin.	Kinashi & Shinagi, 1987; Kinashi *et al.*, 1987
pSCL	*Streptomyces clavuligerus*	12	encodes β-lactam antibiotics; has ITRs and TPs	Keen *et al.*, 1988
SCP1	*Streptomyces coelicolor*	350	ITRs 70Kb; TPs; encodes methylenomycin antibiotics	Kinashi *et al.*, 1987

	Borrelia bergdoferi & B. hermsii (spirochates)	16, 49 & 29	18 bp ITR; dont have TPs - telomeres or ss hairpin loops; present in all strains, 16-200 kb long. Involved in avoidance host immune response.	Plasterk *et al.*, 1985; Hinnebusch *et al.*, 1990; Hinnebusch & Barbour, 1991.
Phages.				
HB3, HB623 & HB746	*Streptococcus pneumonia*	40 (genome)	All have same 23 kDa TP	Romero *et al.*, 1990
Nf, PZA, ϕ15, ϕ29 etc	*Bacillus* phages	18-20 (genome)	have ITRs & TPs; TP bound via serine OH in all cases; TP Nf 66% homolg to ø29 & PZA; encodes Dpols	Paces *et al.*, 1985; Leavitt & Ito, 1987; see section 1.6
PRD1 (& PR4, 5, L17 & PR722)	Gram negative bacteria	14.7 (genome)	PRD1 TP 29.3 kDa, dGTP linked via tyrosine, no sequence homologies to other TPs, ITRs 110-111 bp; Dp 65 kDa Dpol.	Bamford *et al.*, 1983; Bamford & Mindich, 1984; Savilahti & Bamford, 1986 & 1987; Hsieh *et al.*, 1987; Savilahti *et al.*, 1989; Gerendasy & Ito, 1990
Cp-1, 5, 7& 9	*Streptococcus pneumoniae*	18 (genome)	All have TPs of 28 kDa & ITRs (236, 343, 347 & ?bp's); Terminal nucleotide dATP attached *via* threonine, encodes Dpol.	Lopez *et al.*, 1984; Escarmis *et al.*, 1985; Garcia *et al.*, 1983 & 1986.
Viruses.				
Ad	mammals & birds	36 (genome)	TPs, encode own Dpol, ITR (length depends on serotype)	(see section 1.6)

1.6 Protein Primed DNA Replication in Adenovirus and Bacteriophage ø29.

The eukaryotic Ads have a linear 36 kb ds DNA genome and replicate in the host nucleus, using both host- and self-encoded proteins. In contrast, ø29 is one of a number of linear ds DNA bacteriophages of the gram positive bacterium *Bacillus subtilis*. The genome is 19.3 kb long, encoding all the functions necessary for complete replication of the bacteriophage DNA. The genomes of both Ad and ø29 were originally isolated from viral particles as circular molecules and concatemers, which were converted to unit length monomers upon proteolysis (Robinson *et al.*, 1973). Such treatment of the nucleoprotein complex also abolished the transfectivity of the genome (Hirokawa, 1972; Sharp *et al.*, 1976). Furthermore, the resistance of the genomes to 5'-3' (but not 3'-5') exonuclease digestion, the inability to phosphorylate the 5' DNA termini with polynucleotide kinase and the aberrant migration of terminal restriction enzyme fragments in agarose gels (unless pre-treated with protease or SDS) suggested that the 5' end of the genomes were blocked by a protein molecule (Ito, 1978). A 55 kDa protein molecule was subsequently found to be covalently attached to the 5' end of each Ad DNA strand (Rekosh, 1977) and a 28 kDa protein at the 5' termini of the ø29 genome (Yehle, 1978; Salas *et al.*, 1978; Harding & Ito, 1980). The genomes are bordered by ITR sequences of DNA to which the TPs are attached and the 5'-most nt in Ad is always dCMP (Carusi, 1977) and dATP in ø29 (Hermoso & Salas, 1980). The precise length of the Ad ITRs is dependent upon the viral serotype, but ranges from 102-162 bp (Arrand & Roberts, 1979; Tolun *et al.*, 1979; Shinagawa & Padmanabhan, 1979). The ø29 ITR is 6 bp long (Escarmis & Salas, 1982; Yoshikawa & Ito, 1981) and bears no sequence similarity to those of Ad.

1.6.1 Minimal Requirements for DNA Replication.

The elucidation of the mechanism and factors involved in Ad and ø29 DNA replication (reviewed by Salas, 1983; Tamanoi & Stillman, 1983; Sussenbach & Vliet, 1983; Friefeld *et al.*, 1984; Salas, 1988; Salas, 1991) has largely been the result of the development of *in vitro* DNA replication systems. The first such Ad system was developed in 1977, but a more refined system was later developed (Challberg & Kelly, 1979a & b) which demonstrated the involvement of both host and viral factors, and simplified further fractionation and purification of replicatory proteins. A minimal ø29 *in vitro* DNA replication system was established in 1985 (Blanco & Salas, 1985b). The charging of the TP (or pTP) with the terminal nt during

initiation complex formation (ICF) provided a specific assay for the purification of proteins involved in ICF, by the incorporation of [^{32}P]-dCTP or dATP (Shih *et al.*, 1982; Watabe *et al.*, 1982; Challberg *et al.*, 1982; Penalva & Salas, 1982). This subsequently led to the development of a second assay for factors involved in elongation of the IC. Replicative DNA synthesis (in the presence of TPs) could be distinguished from low level non-replicative repair-like synthesis (in the absence of TPs) by the selective [^{3}H]-thymidine pulse labelling of the terminal DNA fragments during short incubations (Lechner & Kelly, 1977; Challberg & Kelly, 1979 a, b; Harding & Ito, 1980). Newly synthesised DNA strands were also weakly bonded to the template and associated with TP; in Ad with the 80 kDa pTP (Challberg *et al.*, 1980). The minimal requirements for complete replication of the Ad, ø29 and PRD1 genomes are summarised in Table 1.5.

1.6.2 Viral Template Requirements.

The minimal template requirements for both Ad and ø29 were elucidated by analysis of the replicatory competence of deletion mutants. Deletion mapping and sequence analysis revealed that the Ad ITR was composed of two main regions (Stillman *et al.*, 1982b): an A+T rich domain within the terminal 51 bp, which supported ICF *in vitro* to 90% of the level attained with full length template; a G+C rich region making up the remaining 50-100 bp, the function of which is unknown but appears to be completely dispensable. More detailed mutational analysis subsequently demonstrated the presence of essential conserved domains within the A+T rich region (Table 1.6; Challberg & Kelly, 1979a & b; Guggenheimer *et al.*, 1984a & b; Rawlins *et al.*, 1984; Hay, 1985b; Wang & Pearson, 1985; Rosenfeld *et al.*, 1987; Wides *et al.*,1987; Hay *et al.*, 1988; Dobbs *et al.*, 1990).

Base substitutions, deletions or insertions within these domains that reduced the efficiency of ICF therefore defined the specific nt and spacings critical for protein binding. Domains X and A constitute the minimal *ori*; domains B and C are auxiliary sequences which increase the efficiency of initiation (ie enhancer regions), but cannot initiate DNA replication on their own. Similarly, deletion and mutational analysis of the SV40 genome demonstrated that the 64 bp minimal *ori* contained three distinct functional domains, separated by spacer regions. Single base substitutions and deletions within these domains drastically reduced replicative

Table 1.5 The Minimal *In Vitro* Requirements for Complete DNA Replication in Ad, ø29 & PRD1. The components outlined below are the only factors required for full length genome replication *in vitro*. Thus both ø29 and PRD1 encode all the proteins required to completely replicate their genome. However, Ad also requires three host-encoded proteins. See text and Table 1.7 for further details.

	Ad	**ø29**	**PRD1**
Virally encoded	Ad DNA-TP	ø29 DNA-TP	PRD1 DNA-TP
	pTP	TP (p3)	TP (p8)
	Dpol	Dpol (p2)	Dp (p1)
	DBP	ds DBP (p6)	
		ss DBP (p5)	
Host encoded	NF I, II & III (see Table 1.7)		
	all four dNTPs	all four dNTPs	all four dNTPs
Others	Mg^{2+}	Mg^{2+} & NH_4^+	Mg^{2+} & NH_4^+

Table 1.6 The Minimal Template Requirements for Ad DNA Replication: Components of the ITR. The main features of the Ad *ori*, contained within the genomic ITR, are outlined below.

ITR region	Domain	Position (nt)	Function	Essential in DNA replication?	References.
A+T-rich		1-51	*ori*	Yes	
	X	1-8	Spacer region. Exact nt sequence not conserved; some nt changes tolerated. Length (4-8 nt) is essential.	Yes	Bergan *et al.*, 1983
	A	9-18	Dp-pTP binding site. Sequence conserved in all serotypes. First 18 bp constitute minimal *ori*, supports initiation *in vitro* up to 3% level of full length genome.	Yes	Tolun *et al.*, 1979; Tamanoi & Stillman, 1983a & b; Rosenfeld *et al.*, 1987; Temperley *et al.*, 1992.
	B	19-40	NF I binding site. Conserved in most (but not all) serotypes. When added to first 18 bp, boosts efficiency ICF 10-fold *in vitro*. Both sequence & length critical.	Yes, when present.	Rawlins *et al.*, 1984; Hay *et al.*, 1988.
	C	41-51	NF III binding site. Conserved in all serotypes. When present, in addition to first 40 bp, increases efficiency ICF further 3-fold.	No	Pruijn *et al.*, 1986 & 1987; O'Neill & Kelly, 1988.
G+C rich region		52-160 (depending on serotype)	unknown	No	

efficiency, as did insertions or deletions within the spacer regions. Like Ad, the minimal SV40 *ori* is also flanked by enhancer sequences which have been implicated in transcriptional regulation (reviewed by Kelly *et al.*, 1988).

Circular vectors containing the extreme terminal regions of the ø29 genome, or linearised derivatives of these circular plasmids maintaining the internalisation of the ø29 sequences, were not templates for ICF. However, linearisation resulting in exposure of the ø29 sequences at the molecular termini allowed template activity, to about 15% of that observed with deproteinised ø29 genome (Gutierrez *et al.*, 1986a & b) and was consistent with previous observations made with the Ad system. Deletion analysis of the extreme termini subsequently defined the minimal ø29 *ori* to lie within the terminal 12 bp. SDM at the second or third nt positions completely abolished template activity, although some alterations at positions 4-12 were permitted. Terminal ø29 DNA fragments as short as 10 bp have been shown to support a limited degree of ICF, but efficiency increases with length of the template. DNA fragments of 10, 26, 42, 73 and 175 bp support initiation to 14, 27, 32, 51 and 74% of the level observed with full length genome, respectively (Garcia *et al.*, 1984). The ligation of non-specific DNA to the terminal 10 bp fragment greatly increased ICF efficiency, suggesting that fragment size and not sequence was the critical factor. Although proteolysis of templates reduced ICF efficiency *in vitro* by about 90%, replication still initiated specifically at the termini (Matsumoto *et al.*, 1984; Grimes & Anderson, 1989). Experiments using ø29 templates containing both, one or no TP molecules demonstrated that, in contrast to *in vitro* results, the presence of TPs at both termini was absolutely essential for replication *in vivo* (Escarmis *et al.*, 1989).

1.6.3 Virally Encoded Protein Requirements.

Initial attempts to purify the proteins involved in Ad and ø29 DNA replication were aided by the development of an *in vitro* complementation assay. This was based on the isolation of conditional lethal temperature sensitive (*ts*) replicative mutants, and their ability to support ICF and its subsequent extension *in vitro*. Mutants were characterised on the basis of the protein fractions which restored replicatory competence, at restrictive temperature, to extracts prepared from cells infected with these mutants. This strategy initially led to the

identification and purification of two Ad encoded proteins (Ostrove *et al.*, 1983), the DBP and the pTP. The latter was found to be tightly associated with another protein, which was subsequently identified as a Dpol. Similar experiments with ø29 *ts* mutants revealed that the products of genes 2 and 3 were involved in ICF whereas those of genes 5 and 6 were involved in late elongation (Mellado *et al*, 1980). Recombinant ø29 genomes containing only the replicative genes 1,2,3,5 and 6 were subsequently found to be able to replicate autonomously; defects in any of these genes could be complemented in *trans* (Escarmis *et al.*, 1989). The minimal *in vitro* ø29 DNA replication system requires only two phage-encoded proteins, p2 (the Dpol) and p3 (the TP) although proteins p5 and p6 greatly enhance the efficiency of replication.

1.6.3 i The Terminal Protein.

The Ad TP is synthesised in the form of an 80 kDa precursor (Challberg *et al.*, 1980). This is active in ICF and is processed late in infection to the mature 55 kDa form by a virally encoded protease (Ikeda *et al.*, 1982). The structural relationship between the pTP and the TP was demonstrated by peptide mapping which revealed extensive homology and by mutations in the protease gene which led to accumulation of the precursor form (Challberg & Kelly, 1981).

The precise role of the mature TP in Ad DNA replication remains obscure, and the significance of the interaction between the template bound TP and the priming TP in linear plasmid replication is not clear. Although the pTP:Dpol heterodimer is known to interact specifically with sequences in the Ad *ori*, specific interaction between the two TP molecules is thought to increase the efficiency of ICF. The genome-linked TP is not essential (Tamanoi & Stillman, 1982) but proteolysis reduces ICF efficiency 10-fold *in vitro* (Bergan *et al.*, 1983) and 100-fold *in vivo* (Hay *et al.*, 1984). In the absence of the TP DNA replication *in vitro* gives rise to internal initiations, close to the terminal nt. The TP may therefore ensure initiation at the correct position, or may simply protect the termini against nucleases and recombination. Analysis of the binding of purified pTP-Dpol heterodimer to a series of deletion mutants, demonstrated that domain A (see Table 1.6) of the Ad *ori* was crucial in DNA binding (Rijnders *et al.*, 1983). Comparative binding studies using Dpol and pTP-Dpol suggested that the pTP moiety was responsible for template binding (Rijnders *et al.*,1983; Chen *et al.*, 1990

Kobayashi *et al.*, 1991; Temperley & Hay, 1992). A short peptide sequence (Arg-Gly-Asp; RGD) within the ø29 TP, close to the priming serine residue, and a corresponding receptor site, have recently been identified (Kobayashi *et al.*, 1989 & 1991). This sequence is also found in some cell adhesive proteins and may constitute the recognition sequence for the priming TP. Synthetic RGD peptides were found to inhibit ICF *in vitro* by specifically binding the template bound TP. Competition studies with the synthetic peptide and priming TP:Dpol demonstrated that increasing RGD peptide concentrations gradually inhibited ICF, until it was completely abolished. However, the RGD peptide did not bind the TP:Dpol, suggesting the receptor site was not exposed. The addition of the first dNTP was subsequently found to be sufficient to expose the receptor site and also reduce the affinity of the TP for the Dpol. Therefore the interaction of the TPs presumably guides the Dpol to the correct position for ICF, the binding of the first dNTP releasing the Dpol for subsequent elongation. Interaction of the template-bound and priming TP may therefore have a greater significance in ø29 than Ad, the specific nt elements being the critical components in the latter case.

Isolation of ø29 *ts3* mutants thermolabile for transfection at the restrictive temperature, and tryptic peptide analysis, demonstrated that the TP (p3) was phage encoded (Salas *et al.*, 1978; Harding & Ito, 1980). Nuclease and protease digestion of the ø29 genome, and alkaline hydrolysis of the resultant nucleotidylpeptide released the terminal nt, dAMP (Hermoso & Salas, 1980); the susceptibility of the linkage suggesting that the DNA was bound to a serine residue (as threonine and tyrosine mediated phosphodiester bonds are not susceptible to alkaline hydrolysis). Proteolysis of the TP, isolation of the peptide bound to the DNA and amino acid sequencing revealed that the genome was linked to serine^{232}, close to the C-terminus of the peptide (Hermoso *et al.*, 1985). Studies with ø29 ICF have also indicated that the base selection for charging of the TP is dependent only on the terminal nt in the complementary strand; in the absence of the correct template any of the four dNTPs are added (Blanco *et al.*, 1992b). Isolation of the native TP was facilitated by the generation of antibodies, raised against TP partially purified from phage particles, and to a synthetic peptide corresponding to the C-terminal eight residues of the protein (Shih *et al.*, 1984). The

immunopurified fraction was active in ø29 DNA replication *in vitro*, and ICF was specifically inhibited by these antibodies.

The Ad pTP and TP are linked to the terminal dCMP through a common region, by a phosphodiester bond to serine^{580} (Challberg *et al.*, 1980; Desidero & Kelly, 1981). Replacement of this residue with threonine completely abolished priming activity, despite the fact that threonine also has a potentially priming β-OH-group. This highlights the spatial constraints which must exist in the positioning of the amino acid side chains at the site of DNA attachment. A similar effect was observed following SDM of the recombinant ø29 TP (Garmendia *et al.*, 1988). All such mutant proteins were able to interact with the Dpol and template, confirming the specific involvement of these residues with the priming function.

The TPs of both Ad and ø29 have been over-produced (in *E. coli* and Vaccinia, and *E. coli* respectively), and the recombinant proteins fully complemented *ts* mutants *in vitro* (Garcia *et al.*, 1983). There have also been limited attempts to define the functional domains of the ø29 TP by deletion analysis. This demonstrated that residues within the extreme N- and C-termini were essential for ICF activity (Zaballos *et al.*, 1986, Pettit *et al.*, 1989, Zaballos *et al.*, 1989a & b; Fredman *et al.*, 1991).

1.6.3 ii The DNA Polymerase.

Cellular Dpols, interacting with the virally encoded TP, were initially thought to be responsible for Ad DNA replication. However, several *ts* DNA replicatory mutants were mapped to the Ad genome and were unable to synthesise viral DNA at the restrictive temperature *in vivo* (Friefeld *et al.*, 1983). These could be complemented by a cytoplasmic fraction from infected cells which contained a Dpol and 3'-5' exonuclease activity (Stillman *et al.*, 1982a; Field *et al.*, 1984). This activity could not be isolated from uninfected cells, or complemented by a wide range of Dpols (Field *et al.*, 1984). The Dpol was subsequently found to be tightly associated, and co-purified with, the pTP (Stillman *et al.*, 1982a; Enomoto *et al.*, 1981; Lichy *et al.*, 1982). Separation and isolation of the Dpol and pTP, and assay for their ability to bind dCTP during ICF revealed that 'charging' was only observed when the two fractions were combined. This suggested that the Dpol exists and catalyses the formation of the IC *in vivo* as a heterodimer with the pTP. The Ad Dpol was also the only Dpol which

could utilise the pTP as a primer for DNA synthesis (Field *et al.* 1984). The Ad Dpol also appeared unusual in that it could translocate through long stretches of duplex DNA in the absence of ATP hydrolysis, although elongation was stimulated 3 to 10 fold by ATP (Lindenbaum *et al.*, 1986). In fact, none of the other enzymes required for Ad DNA replication could unwind duplex DNA under the conditions used to detect the T7 helicase, suggesting that the Dpol may also have an intrinsic helicase activity.

The Ad Dpol gene has been over-expressed *in vitro* (Hassin *et al.*, 1986; Shu *et al.*, 1987), in *E. coli* (Friefeld *et al.*, 1985; Rekosh *et al.*, 1985), Vaccinia (Stunnenberg *et al.*, 1988; Nakano *et al.*, 1991; Joung *et al.*, 1991) and baculovirus (Watson & Hay, 1990). A six residue synthetic peptide, corresponding to the extreme C-terminus of the protein, was also used to raise antisera (Rekosh *et al.*, 1985). Although these antibodies recognised the denatured protein, they did not inhibit replication *in vitro*. This may have been because the native conformation of the protein was not recognised (see Chapter 5) or the antibodies bound to part of the protein not involved with its enzymic activity.

The DBP, pTP and Dpol genes are contiguous on the Ad genome and share several non-contiguous mRNA leader sequences which are differentially spliced to yield the mature mRNA (Stillman & Lewis, 1981; Stillman *et al.*, 1982a). The pTP and Dpol share at least two exons; that located the furthest upstream being only 9 bp long but encoding a tripeptide including the initiating ATG. This sequence is essential for both pTP and Dpol activity, although it has virtually no effect on the size or amount of the protein produced (Shu *et al.*, 1987, Pettit *et al.*, 1988). Omission of this exon from recombinant Dpol, however, led to the production of a protein which was enzymically inactive, could not complement *ts2* mutant cell extracts *in vitro* and was no longer targeted to the nucleus. The addition of the upstream exons, resulted in restoration of nuclear localisation and ICF activity *in vitro* (Stunnenberg *et al.*, 1988). This also confirmed earlier genetic evidence for the occurrence of Dpol mutations outside the main ORF. The minimal coding capacity for an active Ad Dpol was further defined by functional analysis of the enzyme produced by recombinant baculovirus (Watson & Hay, 1990). This revealed that the initiating ATG within the tripeptide was the only absolute requirement for activity, the nature of the other amino acids being unimportant.

DNA binding domains in proteins are often associated with zinc finger' or 'helix-turn-helix' motifs (which usually consist of CXXC separated by about twelve residues from another CXXC or HXXH). These structures are widespread among eukaryotic DNA binding regulatory proteins. There are two such domains in the Ad Dpol, between residues 228-256 and 1060-1090 and SDM has been used to analyse their function in DNA replication *in vitro* (Joung *et al.*, 1991). Mutations in the downstream motif produced an enzyme which was inactive in most assays, but mutations in the upstream motif produced an enzyme which no longer recognised the *ori* (as shown by gel retardation assays) but still interacted with the other replicatory proteins. This suggested that although the upstream region was not involved in protein-protein interactions, it was involved in DNA binding. Mutations at other sites which resulted in the Dpol only weakly interacting with the pTP, also rendered the protein inactive in replication.

ø29 was found to be capable of replicating in Dpol I deficient *B. subtilis* mutants, in the presence of inhibitors of *B. subtilis* Dpol III, DNA gyrase or Rpol (Blanco *et al.*, 1983; Watabe & Ito, 1983). Elongation, but not ICF, was also inhibited by ddATP. The ø29 encoded Dpol was first identified and isolated by fractionation of infected *B. subtilis* cell extracts (Watabe & Ito, 1983). The Dpol also co-purified with a 3'-5' exonuclease activity (Watabe *et al.*, 1984a & b; Blanco & Salas, 1985a). Isolation of *ts2* mutants and complementation using the purified protein confirmed p2 was essential for ICF (Matsumoto *et al.*, 1983; Blanco *et al.*, 1983). Furthermore, the Dpol isolated from *ts*2 infected *B. subtilis* exhibited a much greater thermolability than the wild type enzyme and also rendered ICF thermolabile. Aphidicolin resistant ø29 mutants also mapped to gene 2, and isolation of the mutant protein revealed resistance to 500 mM aphidicolin (Matsumoto *et al.*, 1986). The ø29 Dpol over-expressed in *E. coli* supported ICF in the purified *in vitro* replication system and complemented *ts*2 mutants (Blanco & Salas, 1984; Blanco *et al.*, 1984). The recombinant protein also exhibited the same properties as those previously reported for the authentic enzyme isolated from ø29-infected *B. subtilis*.

The requirement for only two proteins in the minimal *in vitro* ø29 DNA replication system, probably reflects, in part, the inherently high processivity of the ø29 Dpol. This property is

usually conferred upon the enzyme by accessory factors. The ø29 Dpol is also capable of simultaneously unwinding the parental ds helix and displacing the ss DNA during polymerisation (ie. also appears to contain a helicase activity). In addition, factors which increase the stability of template secondary structure do not seem to affect the Dpol. Using primed M13 DNA, the Dpol could synthesise DNA chains of over 70 kb (Blanco *et al.*, 1989). Full length DNA was, however, only synthesised at a reduced rate, suggesting that other factors were required for optimal efficiency (Blanco & Salas, 1984; Blanco & Salas, 1985b). NH_4^+at 10-20 mM, was subsequently found to strongly stimulate ICF *in vitro* (3-fold) and to increase the rate of elongation about 2-fold (Blanco *et al.*,1987). This has also been observed in the minimal *in vitro* PRD1 DNA replication system (Savilahti *et al.*, 1991) and is thought to result from stabilisation of the Dpol-TP heterodimer, which is not detected in the absence of NH_4^+.

A minimal DNA replication system, using purified proteins has now also been established *in vitro* for the *E. coli* bacteriophage PRD1 (Savilahti *et al.*, 1991). The system does not require host factors (Table 1.5) and DNA synthesis occurs at a rate comparable to that of ø29. Like Ad and ø29, the PRD1 TP and Dpol genes are contiguous, co-ordinately transcribed from the same promoter and both are required for ICF. The Dpol was cloned and over-expressed in *E. coli* and although extracts prepared from these cultures could complement PRD1 Dpol mutants, in DNA replication *in vitro*, the over-expressed protein could not be detected by SDS-PAGE. Manipulation of the upstream and downstream regions of the expression plasmid (primarily the ribosome binding site) resulted in high level expression. Most of the recombinant enzyme, however, was produced in an insoluble form. The small amount of soluble enzyme was subsequently purified, and was fully functional in the minimal system.

1.6.3 iii The Adenovirus DNA Binding Protein.

The Ad DBP was originally isolated from crude HeLa cell nuclear extracts by its ability to bind to ss DNA and to complement *ts* DNA replicative mutants, at the restrictive temperature (Horwitz, 1978; Vliet *et al.*, 1978; Schechter *et al.*, 1980). This subsequently demonstrated that the DBP is not only involved in DNA replication (Kaplan *et al.*, 1979), but also gene expression, host range, host infection and virion assembly. The Ad DBP could not be substituted by *E.*

coli or other eukaryotic DBPs (Cleat & Hay, 1989a). In the absence of Ad DBP the Ad Dpol could only synthesise short DNA chains, suggesting that it conferred processivity on the Ad Dpol (Field *et al.*, 1984; Lindenbaum *et al.*, 1986; Chen *et al.*, 1990), in a similar manner to the accessory factors required by other Dpols (eg. the p44, 45 and 62 accessory proteins of T4 Dpol and the thioredoxin subunit of T7 Dpol). DNA footprinting data revealed that the DBP bound the entire length of the Ad genome, suggesting a lack of sequence specificity. The DBP stimulates ICF by increasing the rate of nuclear factor (NF) I binding (Cleat & Hay, 1989a & b; Stuiver & Vliet, 1990), probably by maintaining the *ori* DNA in an extended configuration and stabilising NF I on the *ori*.

1.6.3 iv. ø29 Proteins p5 and p6.

Purified p5 appears to be a ss DNA binding protein, probably protecting ss ø29 DNA from nuclease digestion and stabilising it during replication (Gutierrez *et al.*, 1991a). The isolation of *ts 5* mutants has demonstrated that p5 is essential for elongation *in vivo*, but not *in vitro*. It stimulates DNA replication *in vitro* 2-3 fold and can be functionally replaced by the DBP of *E. coli*, T4 and Ad (Martin & Salas, 1988; Martin *et al.*, 1989; Gutierrez *et al.*, 1991a & b).

p6 appears to be a ds DNA binding protein and is currently thought to be involved in the coupling of initiation and elongation, and maintenance of the DNA in the correct topology for replication; the active form of the protein being a dimer (Blanco *et al.*, 1986). The addition of pure p6 to the minimal *in vitro* system stimulated ICF and although experiments suggested p6 was not essential for DNA replication *in vitro* , *ts*6 mutants were unable to complete replication of their DNA *in vivo* at restrictive temperature. Deletion analysis has revealed that p6 specifically recognises short regions within both ø29 DNA termini, although they exhibit no sequence similarity. The protein, therefore probably recognises a higher order structural feature (Otero & Salas, 1989). A similar observation has recently been reported for the ORF 10 product of K2 (McNeel & Tamanoi, 1991).

1.7.4 Adenovirus Host Encoded Requirements.

Fractionation of uninfected HeLa nuclear extracts, and analysis of the stimulatory effects of these fractions upon Ad DNA replication *in vitro*, led to the identification of 3 host-encoded proteins, NF I, II and III (Table 1.7). Both NF I and III were purified on the basis of their

Table 1.7 The Host Encoded Protein Requirements for Ad DNA Replication. The protein factors outlined below are required by Ad, in addition to its self-encoded proteins, for complete replication of its genome. See text for further details.

Component	MW (kDa)	Host function	Recognition site in Ad	Viral function
NF I	58	Transcription factor CTF-1	Domain B nt 18-40	Facilitates melting of *ori*, stabilizes binding of Dp-pTP & Ad DBP.
NF II	?	Type 1 topoisomerase (see Table 1.1)	-	Type 1 topoisomerase
NF III	92	Transcription factor OTF-1	Domain C nt 41-51	Facilitates melting of *ori* during ICF.

specific DNA binding capability (Nagata *et al.*, 1982; Leegwater *et al.*, 1985; Diffley & Stillman, 1986, Rosenfeld & Kelly, 1986, Jones *et al.*, 1987; Rosenfeld *et al.*, 1987). NF I and III have also been implicated in cellular and viral transcription, and are now thought to be identical to CTF-1 and OTF-1, respectively (Adhya *et al.*, 1986; Jones *et al.*, 1987; O'Neill *et al.*, 1988; Pruijn *et al.*, 1989). The stimulatory effects on DNA replication, however, occur in the absence of transcription excluding the possibility of transcriptional activation of DNA replication as occurs in some other systems. The NFs are thought to be involved with alteration of the DNA topology at the *ori*, making the DNA more accessible to the viral replicatory proteins, and hence allowing the formation of an open complex, in a similar manner to the SV40 large T-antigen (Stuiver & Vliet, 1990). In contrast to ds Ad DNA, the majority of ss DNA sequences tested were templates for replication and did not exhibit any specific nt requirement (Challberg & Rawlins, 1984). The high sequence specificity may, therefore, be determined by the ds nature of the Ad DNA and its ability to bind cellular proteins which perturb the ds structure.

Deletion mutant analysis and DNase 1 protection studies revealed that NF I dimers bound and protected a palindromic sequence at domain B within the *ori*. Furthermore, the ability of NF I to recognise various mutant *ori* directly correlated with its ability to enhance ICF *in vitro*; mutations that reduced binding, reduced ICF efficiency and defined the important nt. The genes for both NF I and III have been cloned and over-expressed in Vaccinia virus and the recombinant proteins were fully active in the stimulation of DNA replication *in vitro* (Gounari *et al.*, 1990). The stimulation of DNA replication by NF I was strongly dependent upon the presence of DBP (Vries *et al.*, 1985; Stuiver & Vliet, 1990). The DBP appears to increase the binding affinity of NF I for its recognition site, by enhancing the rate of association and decreasing the rate of dissociation. NF I also enhances Ad DNA replication by increasing the stability of the pTP-Dpol complex on the template (Mul & Vliet, 1992).

Highly purified preparations of NF II were found to contain topoisomerase activity, and could be substituted by wide range of eukaryotic type 1 topoisomerases, but not by *E. coli* topoisomerase. Although NF II had no effect upon ICF *in vitro*, it stimulated elongation 15-fold, but only in the presence of NF I. Omission from the purified system led to the

generation of daughter DNA strands only 25-30% of the full length (Ikeda *et al.*, 1981; Nagata *et al.*, 1983). The addition of a type 1 topoisomerase inhibitor to the *in vitro* system immediately halted DNA replication (Schaak *et al.*, 1990a & b). NF II is, therefore, probably involved in resolving complex structures which arise during the translocation of the replication fork.

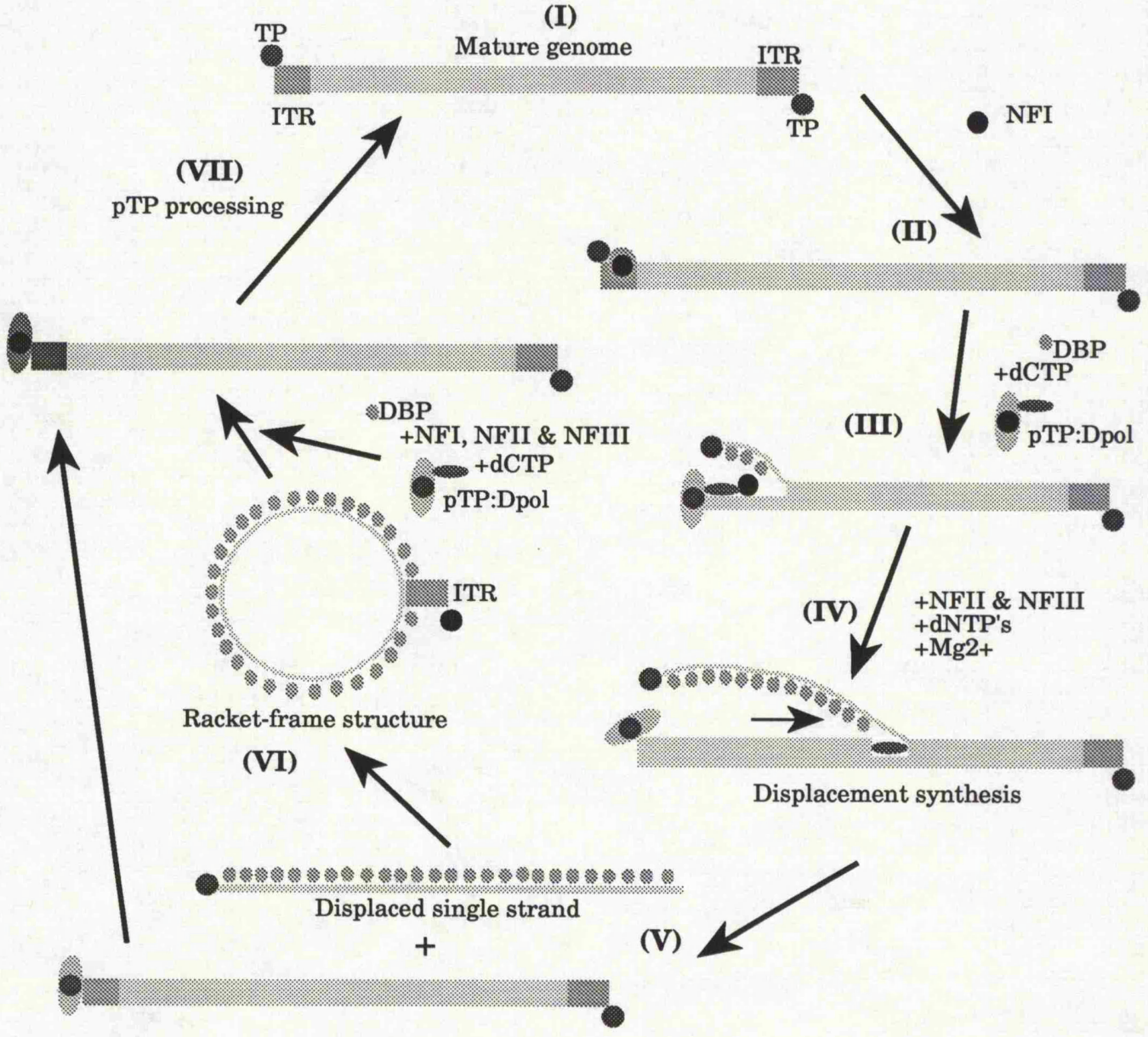
Ad DNA replication *in vitro* was also found to be stimulated 4-6 fold by another nuclear protein, NF III, which DNase 1 footprinting revealed bound domain C within the ITR, adjacent to the NF I binding site (Pruijn *et al.*, 1986 & 1987). SDM subsequently demonstrated that this stimulation was dependent upon NF III binding to its recognition site (O'Neill & Kelly, 1988).

1.6.5 A General Model for Linear Plasmid Replication.

Although the details of this model have been established by work almost exclusively in Ad and ø29 (and its relatives), the number of structural similarities between these and the other linear plasmid systems suggest that the general replicatory mechanism is conserved. The current model for Ad DNA replication is outlined in Figure 1.1.

Biochemical and electron microscopic analysis of replicatory intermediates isolated from Ad and ø29 infected cells suggest that DNA replication proceeds from either end by a strand displacement mechanism (Lechner & Kelly, 1977; Challberg & Kelly, 1979b; Harding & Ito, 1980; Sogo *et al.*, 1982). The frequency of initiation at each end of the genome appears to be about equal and multiple initiation events are common. Two types of displacement intermediates were observed. The first (Figure 1.1, stage IV) consisted of full length ds duplex with one ss branch, of equal length to one of the ds regions. This shows displacement of one strand in progress. The other (stage V) was of full length DNA with one ss and one ds region, and shows complementary strand synthesis of the displaced ss. Similar structures have been reported for the K-plasmids (Fujimura *et al.*, 1988) and the S-plasmids of *Z. mays*. The rate of DNA replication in the purified *in vitro* system, is very similar to that observed *in vivo*. Replication is also initiated specifically at the termini and elongation remains sensitive to high concentrations of aphidicolin. During ICF assays, the (p)TP-dNMP is non-covalently bound to the template and upon addition of the dNTPs this is extended and incorporated into the daughter strand. This was demonstrated by the simultaneous disappearance of the Ad 80

Figure 1.1 The Current Favoured Model forAdenovirus DNA Replication. NFI binds to its recognition sequence (I). This activates the template by localised melting of the extreme DNAtermini, and the ss regions are stabilised by Ad DBP (II). The pTP:Dpol heterodimer binds its now accessible recognition sequence and is stabilised by interaction of the pTP with template bound TP and Dpol with Ad DBP & NFI (III). The Dpol catalyses the covalent charging of the pTP moiety with dCTP. This is aligned with the terminal dGTP of the complementary ss region providing a priming 3' OH group. In the presence of NFII, NFIII, Mg2+ and all four dNTP's, DNA replication proceeds continuously in a 5'-3' direction and the entire genome is replicated from a single priming event (IV). Replication terminates by the complete displacement of the non-template strand, or the meeting of two convergent replication forks (V). If the non-template strand is completely displaced, it may form a pan-handle structure to facilitate subsequent replication (VI). The pTP is processed late in infection (VII).



kDa pTP-dCMP and the appearance of an 88 kDa pTP-26mer, in the presence of ddGTP (Lichy *et al.*, 1981). Similarly, in the presence of ddCTP, the ø29 IC was elongated and subsequent piperidine cleavage released 9 and 12 bp DNA fragments from either end of the genome. This confirmed not only that replication initiates and proceeds from both termini, but that the IC is indeed the first product of replication.

Evidence indicates that, in at least some cases, Ad DNA replication results in displacement of one strand before initiation occurs on the other, resulting in replication of only one strand. The subsequent completion of the displaced ss has been the subject of some speculation. By virtue of the ITRs the displaced ss has the potential to assume a racket-frame (pan-handle) structure which produces a molecular end identical to the parental ds species (the handle). This provides the essential ds recognition site for NF I, and eliminates the necessity for a second replicatory mechanism. The ss molecule is then simply converted to the ds form by complementary strand synthesis. Data generated *in vitro* using Ad mini-chromosomes containing deletions in one of the ITRs indicates a sequence independent requirement for an intact ITR (Stow, 1982). Although mini-chromosomes bearing only one ITR were replicatively defective, deletions at the outer termini of one ITR led to regeneration of the complete sequence in progeny chromosomes. Deletions at the internal end of the ITR, however, were not repaired. This suggests a repair mechanism based on comparison of the two ITRs (Lippe & Graham, 1989). Further evidence supporting panhandle formation has recently emerged (Hu *et al.*, 1992). Some kinetic data, nevertheless, indicates that ICF is 50 times faster but elongation is 2.5 times slower on the displaced strand, suggesting that different mechanisms are operative on each strand. The faster ICF maybe consistent with easier *ori* melting in only short ds DNA regions, and the slower elongation rate may simply reflect the difficulty in removing the Ad DBP. The replication of ø29 is, therefore, thought to differ from Ad in at least one important aspect; the 6 bp ø29 ITRs are probably too short to support the formation of a panhandle-type structure, despite the additional 40% homology within the terminal regions. This, together with the fact that there appears to be no requirement for a NF I-type ds recognition sequences, suggests that replication of the ø29 genome is symmetrical.

1.7.1 The linear plasmids of *Kluyveromyces lactis*.

The dairy yeast *K. lactis* is a lactose fermenting budding yeast which, unlike the brewing yeast *S. cerevisiae*, has an obligate requirement for mitochondrial mediated (ie. oxidative) metabolism. There has, recently, been an increasing awareness of the commercial potential of this organism, due (in part) to its ability to utilise cheap growth substrates such as lactose and whey, a consequence of a chromosomal lactase (β-galactosidase) gene. Furthermore, *K. lactis* appears to be an efficient protein secretor (Berg *et al.*, 1990; Fleer *et al.*, 1991a & b; Romanos *et al.*, 1992). These features have, therefore, increased its suitability as a host for heterologous protein production. For many years *K. lactis* has been used for the large scale production of lactase and single cell protein, thus large scale cultivation of this organism has been extensively studied. More recent developments include the industrial scale production of the milk clotting enzyme prochymosin by Gist-Brocades (Berg *et al.*, 1990) and the successful over-expression of human serum albumin and interleukin-1β (Fleer *et al.*, 1991a & b).

During routine screening of yeast species, certain strains of *K. lactis* were found to contain two unique cytoplasmic linear ds DNA plasmids, pGKL1 (K1) and pGKL2 (K2). The presence of these plasmids strongly correlated with the ability of *K. lactis* strains to inhibit the growth of sensitive strains or species of yeast (Gunge *et al.*, 1981; Panchal *et al.*, 1985). These were the first yeast DNA plasmids discovered to confer a known phenotype. The cytoplasmic location of these plasmids was confirmed by differential centrifugation of osmotically lysed *K. lactis* protoplasts (Stam *et al.*, 1986) and comparisons of killer plasmid-containing rho^0 and rho^+ *S. cerevisiae* strains with a fluorescent DNA-binding dye (Gunge *et al.*, 1982).

The presence of covalently attached TPs on both K-plasmids was first demonstrated using the techniques which had previously been employed with Ad and ø29 (Kikuchi *et al.*, 1984; see Section 1.6). The terminal fragments of K1 and K2 exhibited different mobility shifts when treated with proteolytic enzymes, suggesting the presence of two different size TPs. Radio-iodination of these proteins in purified plasmid preparations, followed by complete digestion of the DNA, allowed subsequent visualisation by SDS-PAGE. This revealed that K1 and

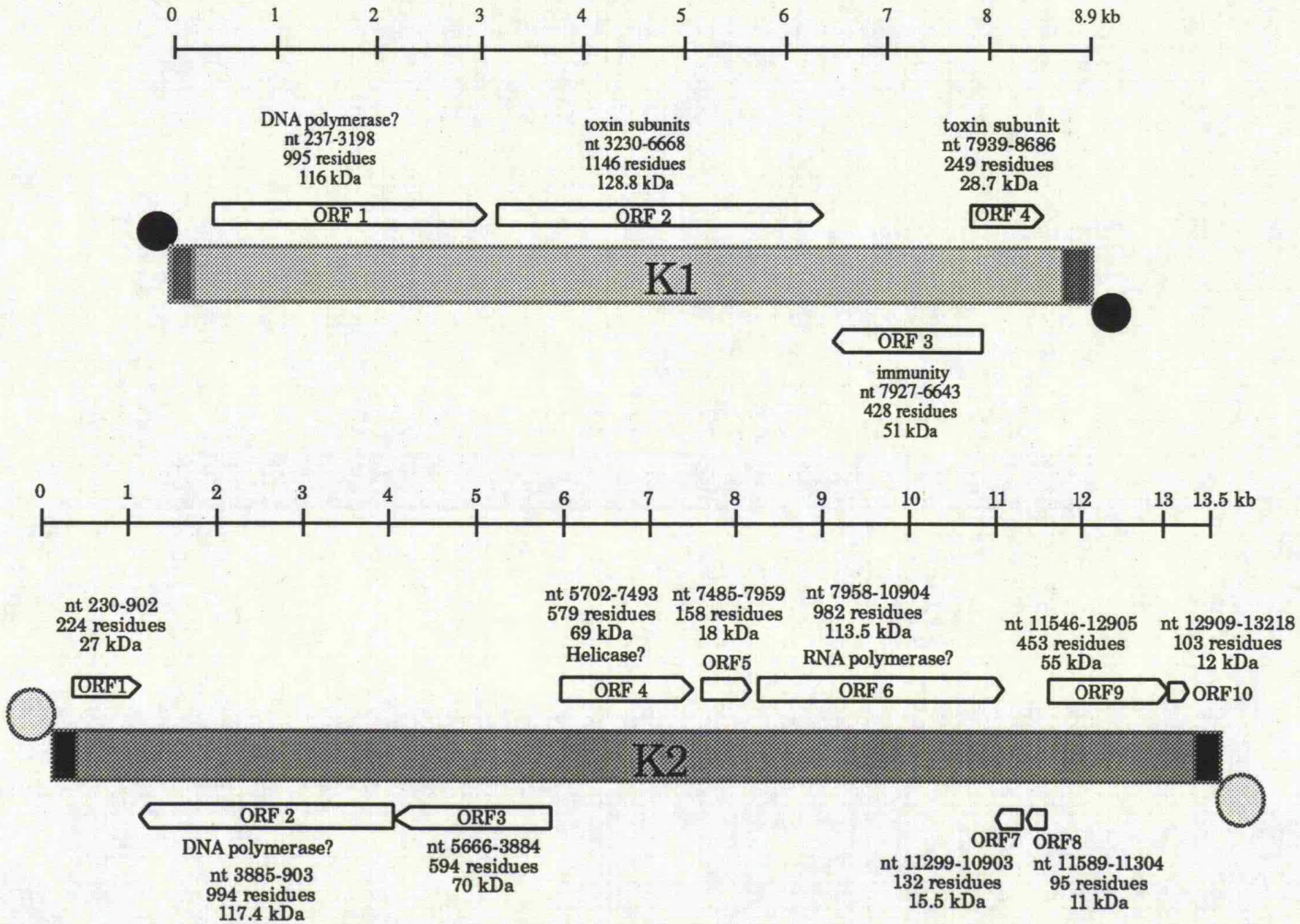
K2 do indeed have different TPs, with molecular weights of approximately 28 and 36 kDa, respectively (Stam *et al.*, 1986).

The K-plasmids are thought to be present at 50-100 copies/cell. Although no firm evidence is available to date, segregation is suspected to result simply from random partitioning during division of the cytoplasm at cytokinesis. If so, copy number may be maintained by continual DNA replication throughout the cell cycle, like mitochondrial DNA. The K-plasmids do not share any sequence similarities with nuclear or mitochondrial DNA, so these replicons are probably not used as carrier for mitotic segregation. Nor is there any evidence for the existence of nuclear genes affecting their maintenance, like the *MAK* (maintenance of killer) genes of the *S. cerevisiae* RNA killer plasmids (Wesolowski *et al.*, 1982b). The K-plasmids also do not contain any sequences resembling the *S. cerevisiae* or *K. lactis CEN* or the 2μ *STB* loci. Both K1 and K2 do contain sequences which confer autonomous replicative ability (ARS) upon nuclear circular DNA plasmids in both *K. lactis* and *S. cerevisiae*. This is, however, probably a consequence of their high A+T content (Sreekrishna *et al.*, 1984); the consensus ARS element also being A+T rich.

1.7.2 K-plasmid Structure.

The complete sequencing of both K1 and K2 (Sor *et al.*, 1983; Hishinuma *et al.*, 1984; Stark *et al.*, 1984; Sor & Fukuhara, 1985; Stark, 1988; Tommasino *et al.*, 1988; Wilson & Meacock, 1988) and analysis of the resultant data has revealed a number of interesting features. Like other extrachromosomal linear ds DNA replicons, K1 and K2 have unique ITRs of 202 and 184 bp (respectively) but these bear no sequence homology to each other (Sor *et al.*, 1983). The 8.9 kb of K1 contains four ORFs; the 13.5 kb of K2 contains at least ten (Figure 1.2). Both K1 and K2 exhibit a remarkable economy of space, over 95% of K1 and 97% of K2 DNA having coding potential and ORFs overlap on opposite strands in both plasmids. As typified by other non-nuclear DNA elements (eg. mitochondrial and chloroplast genomes) the K-plasmids exhibit a high A+T content, over 73% (see Table 1.4), something first implied by analytical density gradient centrifugation. Despite this, the ORFs of both K1 and K2 use the universal, rather than the mitochondrial, genetic code, which supports the idea of a cytoplasmic location for these plasmids. Stark *et al.*, (1984) also noted that the lowest G+C

Figure 1.2 A diagramatic representation of the structure of the linear plasmids, pGKL1 (K1) and pGKL2 (K2), of *Kluyveromyces lactis*. The plan illustrates the presence of different terminal proteins and ITRs, as well as the number and arrangement of ORFs on each plasmid.



content (13.6%) was present in the third base of each codon, where the greatest 'wobble' is permitted. If the codons which show an absolute requirement for a G nt in the third position were taken into account (ATG for methionine and TGG for tryptophan) then this bias was even more pronounced (10.8% G+C). The recent DNA sequencing of the *S. kluyveri* cytoplasmic plasmid pSKL (Kitada & Hishinuma, 1987) has not only revealed an A+T content close to that of K2 (68%), but that the terminal 15 bp are identical to those of K2 (although the ITRs are different lengths) and that pSKL has the same number and arrangement of ORFs as K2.

K1 has been shown to encode the toxin and immunity factors (Stark & Boyd, 1986). Several of the products of K2 ORFs have also been assigned tentative functions. K2 ORF 4 encodes a protein with some sequence similarity to the putative Vaccinia DNA-dependent ATPase, possibly a helicase (Wilson & Meacock, 1988). ORF 10 encodes a small, very basic protein, with a high (25%) lysine and arginine content. The over-expression and analysis of this protein has recently been reported (Tommasino, 1991; McNeel & Tamanoi, 1991) and demonstrated that it binds strongly to K-plasmid DNA. The putative ORF 6 product exhibits three regions of homology with β and β' subunits of prokaryotic and eukaryotic Rpols (Wilson & Meacock, 1988). The three regions correspond to the domains generally believed to be responsible for ribonucleotide binding (ß) and DNA template binding (ß'). Two additional ß' domains have also been detected in K2 ORF 7. ORFs 6 and 7 are, therefore, thought to encode a highly condensed and unique Rpol upon which both plasmids appear to be dependent for their transcription. Both K1 and K2 contain ORFs with the potential to encode Class B Dpols.

Northern analyses of K-plasmid bearing yeast strains revealed 10 K2 transcripts (Fleming & Meacock , unpublished results). Similarly, using the entire K2 sequence as a probe, at least 8 transcripts of 0.4-3.5 kb were detected (Tommasino, 1991). None of the K-plasmid ORFs are preceded by recognisable yeast nuclear promoters, although unusual upstream and downstream sequences have been identified for most of the putative ORFs (Sor & Fukuhara, 1985; Stark *et al.*, 1984; Fleming & Meacock, unpublished results). Primer extension mapping was employed to locate the transcriptional start sites of each of the 10 K2 ORFs (Fleming &

Meacock, unpublished results). All such start sites lay 10-20 bp downstream of a sequence proposed to be the promoter recognised by the K2-encoded Rpol. Similar observations were made with K1 (Romanos & Boyd, 1988). The intervening sequence was found to have the potential to form a stem-loop structure, the start codon being immediately preceded by three A nt which lie on the distal side of this stem. It is for these reasons that the K1 ORF 1 and K2 ORFs 4 and 6 products are thought to initiate at the second in-frame ATG.

1.7.3 K-plasmids in Other Yeasts.

The K-plasmids have been successfully transferred to *S. cerevisiae*, *K. fragilis* and *Candida pseudotropicalis* by protoplast fusion and illegitimate mating, where they are stably maintained and express the killer phenotype (Gunge & Sakaguchi, 1981; Gunge *et al.*, 1982; Sugisaki *et al.*, 1985). Transformation of *S. cerevisiae* led to the discovery of the curious incompatibility of the K-plasmids with *S. cerevisiae* mitochondrial DNA (Gunge & Yamane, 1984). Under certain circumstances *S. cerevisiae* is not dependent upon mitochondria for cell viability and this allowed the study of this phenomena. In the presence of mitochondrial DNA (rho^+ strains) the plasmids exhibited a mitotic instability resulting in segregation of non-killer clones. They were also unstable in *mit⁻ S. cerevisiae* strains (which carry point mutations in the mitochondrial genome) but were stably maintained in some rho^- (carry rearranged or partially deleted mitochondrial genomes) and all rho^0 strains (lack mitochondrial DNA but still contain degenerate mitochondria). The instability appears to be due to the presence of mitochondrial DNA, rather than a product of the DNA. This has been demonstrated by the curing of mitochondrial DNA from rho^+ killer strains; all clones subsequently segregated exhibited the killer phenotype. Furthermore, even chloramphenicol-mediated inhibition of mitochondrial protein synthesis in rho^+ *S. cerevisiae* strains maintained the mitotic instability. These results indicated that competition for rare tRNA species was probably not the basis for the instability, although this could conceivably implicate nuclear encoded mitochondrial gene products. In addition, the transmission of K-plasmids between rho^0 *S. cerevisiae* strains, by protoplast fusion, was found to be unexpectedly low, which was surprising as diploid formation should simply involve the fusion of both cell contents. In contrast to *S. cerevisiae*, *K. lactis* is obligatorily rho^+, so

this incompatibility must have a basis in the differences between the mitochondria of these two species.

The stable maintenance of the K-plasmids has recently been reported to be mating type dependent in *S. cerevisiae*, even in rho^0 strains (Fujimura *et al.*, 1988; Gunge *et al.*, 1990). K-plasmids were stable in rho^0 MATa or MATα homozygous diploids or haploids but unstable in a/α heterozygous diploids. This suggests that the K-plasmids are affected by the *MAT* locus in some way, and hence, that a slightly different mating system exists in *K. lactis*. DNA sequence analysis of the K-plasmids subsequently revealed that a 20 bp motif similar to the recognition site of the a1-α2 haploid specific transcriptional repressor protein occurred three times within K2 coding regions; once within the K2 ORF 2 promoter and twice within ORF 2. The a1-α2 repressor may, therefore, interfere with K2 by binding these sequences and disrupting transcription. The incomplete elimination of K-plasmids even after prolonged subculturing in a/α diploids could be accounted for by titration of the repressor by the high copy number of K2 or inefficient passage of the repressor or plasmids across the nuclear membrane. Alternatively, the nuclear repressor may interact with the cytoplasmic K-plasmids due to leakage of nuclear components.

1.7.4 Mutant K-plasmids

Non-killer derivatives of *K. lactis* have been obtained by UV irradiation of killer strains. Those non-killers which lack just K1 (ie. K1^0 K2$^+$) or both plasmids (ie. K1^{0}K2^0) lack all toxin and immunity subunits, and are therefore sensitive to inhibition. This also demonstrates that K1 is responsible for both the killer and immunity functions. There have not, however, been any reports of K1$^+$ K2^0 strains suggesting that K2 is essential in replication and maintenance of the plasmids.

Non-killer strains carrying defective plasmids have also been isolated. For example, following the transformation of *S. cerevisiae* with the K-plasmids two novel K1-derived plasmids were reported (Kikuchi *et al.*, 1985). These two plasmids F1 (7.8 kb) and F2 (3.9 kb) never segregated from each other or native K1. Restriction mapping and Southern analysis demonstrated that F1 was identical to the left terminus of K1 and had probably arisen due to a 5 kb deletion of the right terminus. F2 had a hairpin-loop structure and

denaturation revealed that it was a perfect head-head ss dimer of F1. Other similar mutant plasmids have also been reported (Wesolowski *et al.*, 1982b; Kitada & Gunge, 1988; Gunge & Kitada, 1988; Wesolowski-Louvel & Fukuhara, 1990; Tanguy-Rougeau *et al.*, 1990). Analysis of these mutant plasmids has also given an insight into the mode of K-plasmid replication; the origin and interconversion of such plasmids being accounted for by an Ad type replicatory mechanism (Figure 1.1). The mutant plasmids are thought to have arisen by endonucleolytic cleavage, followed by exonucleolytic degradation of the unprotected 3' ends. Hairpin formation occurred when internal palindromic (inverted repeat) sequences were exposed, with subsequent DNA extension using the top strand as a template.

1.7.5 Replication of the K-plasmids.

The TP-DNA structures of Ad and ø29 have been shown to play an integral role in the replication of their genomes; the TP supplying the necessary primer for initiation of DNA replication. The number of similarities between Ad, ø29 and the K-plasmids make it likely that the terminal structures of K1 and K2 play a similar role in plasmid replication. Furthermore, the cytoplasmic location of the K-plasmids suggests that DNA replication may rely heavily upon plasmid-encoded functions.

The genes encoding the TPs have yet to be identified, but since all of the ORFs on K1 have been assigned other functions it seems unlikely that K1 encodes its own TP. K2 may, therefore, encode both and should this be the case, it is remarkable that they do not exhibit any sequence similarity. Of the unassigned ORFs on K2 there are three (3, 4 and 9) large enough to encode the TPs. An alternative hypotheses is that a single precursor TP is synthesised (as in Ad) and differentially processed depending upon the plasmid to which it is targeted. If so, the K-plasmids may also encode the necessary protease, as seen in Ad.

The putative Dpols of K1 and K2 share a high degree of sequence similarity (28%) over their C-terminal 700 or so amino acids. A sequence similar to the most highly conserved polymerase motif (YGDTDS) is found within both these ORFs (Jung *et al.*, 1987a; Fukuhara, 1987; Tommasino *et al.*, 1988). The fact that each plasmid has its own TP and Dpol, the lack of homology within the Dpol N-termini (in contrast to that in the C-termini), and the fact that these N-terminal domains have sufficient capacity to encode the TPs, led to the suggestion

that the gene for each Dpol might also encode a TP (Meacock, Pers. Comm.). Similar proposals have since been put forward for the N-terminal domains of the *C. purpurea* pCIK1 Dpol (Oeser & Tudzynski, 1989; Salas, 1991), the *kalilo* Dpol (Chan *et al.*, 1991), the *P. anserina* pAL2-1 Dpol (Hermanns & Osiewacz, 1992) and the Duck Hepatitis B Virus reverse transcriptase (Bosch *et al.*, 1988; Bartenschlager & Schaller, 1988; Wang & Seeger, 1992). The TP of the Duck Hepatitis B virus remains covalently attached to the RNA genome in the virion, and is thought to prime DNA synthesis by the reverse transcriptase after host infection. Interestingly, the size of the kalilo TP (120 kDa) corresponds closely to that of the predicted Dpol. Alternatively, the N-terminal domains of the putative K-plasmid Dpols may function in the targeting to the appropriate plasmid, *via* interaction with the template-bound TP or the DNA template itself. Bearing in mind the compact nature of the K-plasmid genome, the fact that the Dpol and TPs in Ad, ø29 and PRD1 are contiguous, co-ordinately transcribed (and possibly translated) and are often tightly associated *in vivo*, such an arrangement would not be totally unexpected. Although no experimental evidence for any of these models is currently available, circumstantial evidence obtained thus far is consistent with each of the K-plasmids encoding their own Class B Dpol and replicating by a protein primed mechanism, as typified by Ad, ø29 and PRD1.

1.8. Aim of the project.

The description of the K-plasmids in the preceding text illustrates the similarities these elements bear to other 'linear plasmids'. These, therefore, make it highly likely that the K-plasmids replicate in a very similar manner to that used by adenovirus and ø29. However, there is, as yet, no firm evidence available to confirm this.

One of the characteristic features of linear plasmids is their capacity to encode their own highly specific Class B Dpols. Two such genes have been identified in the K-plasmids; ORF1 in K1 and ORF2 in K2. Consequently these are thought to be responsible for the replication of the K-plasmid genomes. However, these genes have only been defined as such on the basis of similarities in the deduced amino acid sequences to other members of the Class B family of Dpols. Although both these genes are known to be transcribed, no biochemical evidence exists to support this theory.

Within the context of the overall aim of work within our laboratory, being to establish a minimal cell free system which was capable of replicating the K-plasmids, the goal of this particular project was to characterise the putative Dpol gene products. With this in mind, the most convenient way of obtaining relatively large quantities of pure protein was initially thought to be by over-expressing the genes in *E. coli.*

CHAPTER 2.

MATERIALS AND METHODS.

2.1. GROWTH MEDIA.

2.1.1 Complex Growth Media.

E. coli strains were grown in Luria broth (LB) containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract, 5 g/l NaCl, pH adjusted to 7.2. When required, LB was solidified with 15 g/l Difco purified agar. All media was sterilised, unless otherwise stated, by autoclaving at 121°C, 15 psi.

K. lactis was grown in Yeast peptone dextrose (YPD) broth which consisted of 10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose. This media was solidified, when required, as above.

2.1.2 Minimal Growth Media.

E. coli were grown in M9 minimal medium. This was made up and autoclaved as two separate concentrated stocks and diluted as required: 10x M9 salts (210 mM Na_2HPO_4, 220 mM KH_2PO_4, 86 mM NaCl, 187 mM NH_4Cl) and 100x M9 additive (0.1 M $MgSO_4$, 0.01 M $CaCl_2$). Glucose was added, as a growth substrate, to 20 g/l. Solid medium was obtained by the addition of purified agar, as above.

In certain circumstances amino acid supplements were added to the *E. coli* minimal medium. A 5 g/l stock solution, containing an equimolar mixture of 18 amino acids (excluding cysteine and methionine to prevent interference with [^{35}S]-methionine incorporation) was used in the medium at final concentration of 0.05 g/l.

K. lactis was grown in semi-defined (SD) media. This was composed of 20 g/l glucose, 6.7 g/l of Difco Yeast Nitrogen Base (containing $(NH_4)_2SO_4$ but not amino acids) and was solidified, when required, as above.

Specific amino acids were also added to the minimal yeast media to complement the auxotrophic strains being used. L-lysine and L-arginine were stored at 4°C as 5 g/l stocks and used in the media at a final concentration of 25 mg/l.

2.1.3 Antibiotics.

Selection of plasmid bearing *E. coli* strains was accomplished by incorporation of either ampicillin or kanamycin (50 g/l stocks, stored at -20°C) in the growth medium to a final concentration of 0.1 g/l.

2.1.4 Other Additions.

Recombinant plasmids in *E. coli* were also sometimes selected by insertional inactivation of the *lacZ* gene product. Both the gratuitous inducer IPTG and the chromogenic substrate X-gal were made up as 50 g/l stocks and used in the selective media at a final concentration of 50 µg/l. The former stock was stored at -20°C, but the latter was made up fresh in dimethylformamide, immediately prior to use.

2.2 BACTERIAL AND YEAST STRAINS AND PLASMIDS.

All *E. coli* and *K. lactis* strains were provided by Dr. P. Meacock (Leicester Biocentre, Department of Genetics, Leicester University) except *K. lactis* JK1 (by Mr. S. Soond, Leicester Biocentre). All *E. coli* (except N4830-1, which was grown at 30°C) were grown at 37°C and all *K. lactis* at 30°C in an orbital shaker at approximately 200 rpm, unless otherwise stated. All strains were maintained on solid media and were restreaked at trimonthly intervals.

Plasmids pKLK111, 201, 203 were provided by Dr. Meacock . pKLK111 consisted of 8.75 kb of K1 cloned into pUC12. pKLK201 contained 5.5 kb of K2 cloned into pUC12. Plasmids pET3a:*myo*I and pET3a:*sup*45 were supplied by Dr. M. Pocklington (Department of Genetics, Leicester University). Plasmid pMS11 was supplied by Mr. S. Soond.

Strain Genotypes.

E. coli NM522 (Gough & Murray, 1983).

Δ(lac proAB) thi supE Δ(hsdMS-mcrB)5 r_k-m_k- *mcrB/ F' proAB*$^+$ *lacI*^{Q} *lacZ Δm15*

E. coli BL21(DE3) (Rosenberg *et al.*, 1987).

F$^-$ *hsdS gal (λcIts857 ind1 Sam7 nin5 lacUV5-T7 gene 1)*

E. coli JA221 (Clarke & Carbon, 1978).

F$^-$ *hsdR trpE 5 leuB6 recA lacY*

E. coli JM105 (Pharmacia Biotechnology).

F' [traD36 lacI^Q *proAB Δ(lacZ)M15] thi Δ(lac-pro) rpsL Str*^{r} *endA sbcB15 hsdR4*

E. coli N4830-1 (Gottesman *et al.*, 1980).

F⁻, Sup0 His⁻ Ilv⁻ GalKΔ8 (ΔchlD-Pgl) [ΔBamN+ cI857 ΔH1]+

K. lactis MG1/2 (Wesolowski-Louvel *et al.*, 1988).

MAT α, *uraA, argA, lysA*, [K1, K2]+ pKD1+

K. lactis JK1 (Soond, Leicester Biocentre).

MAT α, *trp1, argA, lysA*, [K1, K2]+ pKD1+

2.3 MANIPULATIONS WITH BACTERIA AND YEAST.

2.3.1 Measurement of *E. coli* Cell Culture Densities.

Cell densities were estimated by their absorbance at 595 nm, in a plastic 1 ml cuvette (1 cm path length) in a Pye Unicam spectrophotometer. Readings were referenced to the media in which the bacteria were grown (M9 or LB).

2.3.2 Preparation of Competent *E. coli.*

E. coli were made capable of taking up naked plasmid DNA (ie. competent) by the method of Hanahan (1983). A bacterial colony from an agar plate was inoculated into 100 ml of LB and grown overnight. Cells were sub-cultured (1%) into 100 ml of fresh media and grown to mid exponential phase (an OD_{595nm} of about 0.5, taking 2-3 h). The culture was then chilled on ice for 5 min and the cells pelletted (6,000 rpm, 5 min, 4°C), resuspended in 20 ml of transformation buffer 1 (30 mM K acetate, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% v/v glycerol, pH adjusted to 5.8 with 0.2 M acetic acid; filter sterilised and stored at -20°C) and incubated on ice for 5 min. Cells were recentrifuged, resuspended in 2 ml of transformation buffer 2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% w/v glycerol, pH adjusted to 6.5 with KOH, filter sterilised and stored as 2 ml aliquots at -20°C) and incubated on ice for a further 15 min. Aliquots (200 µl) of the culture were then dispensed into pre-chilled Eppendorf tubes, snap-frozen using dry ice-IMS or liquid nitrogen and stored at -70°C, for up to 3 months.

2.3.3 Transformation of Competent *E. coli.*

Cells were thawed at room temperature, 1-2 μg of plasmid DNA was added and the tubes incubated on ice for 40 min. The cells were briefly heat shocked (42°C, 2 min in a water bath) and returned to ice for a further 10 min. LB (800 μl) was added and the cells incubated (45 min, 37°C; except with *E. coli* N4830-1 which was grown at 30°C) with gentle shaking to allow induction and synthesis of plasmid encoded antibiotic inactivating enzymes. Transformed cells (200 μl) were spread on selective media plates and incubated overnight.

2.3.4 Measurement of *K. lactis* Cell Density.

K. lactis was grown in SD medium supplemented with the necessary growth requirements. A 15 μl aliquot of cell suspension was pipetted onto a Neubauer Haemocytometer (depth 0.1 mm, 1/400 mm^2) and the culture was allowed to diffuse and settle under the coverslip. The haemocytometer grid was composed of 25 large squares (defined by triple borders) and each was further divided into 16 smaller squares. The number of yeast cells in a random selection of the 25 large squares was counted and the average number of cells per large square calculated. Each cell in the smallest squares was equivalent to about $4x10^6$ cells/ml, therefore the approximate cell density could be calculated.

2.3.5 Preparation of Competent *K. lactis.*

K. lactis cells were made competent using the two methods outlined below.

Method I (Chen *et al.*, 1992). This combines making the cells competent with the transformation procedure. *K.lactis* was grown for 18-36 h in 10 ml of minimal media. Cells were gently centrifuged (3,000 rpm, 5 min, room temperature) and resuspended in 1 ml of transformation buffer (0.2 M Li acetate, 400 g/l PEG_{3350}, 100 mM DTT). Aliquots of 100 μl were transformed with 1-2 μg of DNA plus 50 μg of single stranded carrier DNA (salmon sperm DNA). Each transformation was incubated (45°C, 50 min), gently plated directly onto selective media and incubated for 7 to 10 days, at 30°C

Method II (Mount *et al.*, 1992). *K. lactis* were grown overnight in 10 ml of minimal medium, and sub-cultured into 50 ml of fresh medium, to a final cell density of about $2.5x10^6$ cells/ml. The culture was incubated for a further 3-5 h, until the cell density

attained about $1x10^7$ cells/ml, then gently pelleted (as above), washed in 10 ml of TTE (10 mM Tris-HCl pH 7.6, 0.1 mM EDTA) and resuspended in 10 ml of 0.1 M Li acetate (in TTE). Cells were then incubated (shaking, 30°C, 1 h), pelleted and resuspended in 1 ml of fresh Li acetate solution as above. Aliquots of 300 µl were then used for each transformation.

2.3.6 Transformation of Competent *K. lactis*.

Competent cells prepared by the method of Mount *et al.*, (1992) were transformed using the following procedure. Up to 10 µg of transforming DNA was added to each 300 µl aliquot of competent cells (along with 40 µg of carrier salmon sperm DNA, to increase the efficiency of transformation). PEG_{4000} (500 g/l, 700 µl) was added, gently mixed and the tubes incubated (30°C, 1 h). Cells were heat shocked (42°C, 5 min) and then 200 µl aliquots were spread on selective media. The remaining 800 µl was gently pelleted, resuspended in 100 µl of water and also spread onto selective media, to obtain a more concentrated plating. Plates were incubated for 7-10 days, until the colonies were sufficiently large to pick.

2.3.7 Rapid Plasmid Preparation from *E. coli*.

DNA solutions for rapid screening of potential recombinant plasmids were originally obtained using the method of Holmes and Quigley (1981) but later the method of Chowdhury (1991) was used. Although, in the latter case, the DNA preparations were very crude, few problems were experienced with restriction digests and transformations.

Method 1 (Holmes and Quigley, 1981). Single colonies of plasmid bearing *E. coli* were inoculated into 5 ml of selective media and grown overnight. Cultures were harvested (5,000 rpm, 10 min, room temperature) and resuspended in 350 µl of STET (80 g/l sucrose, 5% v/v Triton X-100, 50 mM EDTA, 50 mM Tris-HCl pH 8.0) and transferred to Eppendorf tubes. Aqueous lysozyme (10 g/l, 25 µl) was added, the tubes incubated (37°C, 5 min) and then transferred to a PEG bath (at 105°C) for 90 s. The lysed cells were centrifuged (11,000 rpm, 10 min, 4°C), the pellet removed with a toothpick and discarded. An equal volume of isopropanol was added to the supernatant, the tubes were placed in a dry ice-IMS bath for 10 min, centrifuged (as above) and the supernatant discarded. The tubes

were inverted on a paper towel and allowed to dry at room temperature. The pellet was resuspended in 150 µl of TE, phenol extracted, ethanol precipitated and the DNA was finally resuspended in 50 µl of TE; 5-10 µl was used for each digestion.

Method II (Chowdhury, 1991). Single colonies were inoculated into 1 ml of selective media in an Eppendorf tube and incubated for 5-8 h. The cultures were then directly extracted with 300 µl of phenol:chloroform:isoamyl alcohol, the upper phase containing the DNA was ethanol precipitated and resuspended in 50 µl of TE; 5 µl of the resultant DNA solution was sufficient for most digestions.

2.3.8 Semi-purified Plasmid Preparations from *E. coli.*

Larger amounts of semi-purified plasmid DNA were prepared by a method adapted from that of Birnboim and Doly (1979). *E. coli* was grown overnight in 100 ml of selective media, the cells pelleted (9,000 rpm, 5 min, 4°C), resuspended in 1 ml of ice cold lysis buffer (50 mM glucose, 25 mM Tris-HCl pH 8, 10 mM EDTA, 4 g/l lysozyme) and incubated on ice for 30 min. Freshly prepared alkaline solution (2 ml, 0.2 M NaOH, 10 g/l SDS) was added, mixed gently and returned to the ice for 5 min. Na acetate (1.5 ml, 3 M, pH 5.5) was added, the tubes mixed gently by inversion and incubated on ice for 1 h. Cell debris was removed by centrifugation (10,000 rpm, 10 min, 4°C) and the supernatant ethanol precipitated. The pellet was resuspended in 500 µl of TE, treated with RNase, phenol extracted several times, ethanol precipitated a second time and finally resuspended in 200-400 µl of TE.

2.3.9 Highly Purified Plasmid Preparations from *E. coli.*

Highly purified DNA for plasmid constructions, DNA sequencing or *in vitro* transcription was prepared either by buoyant density ultra-centrifugation in a CsCl gradient or using the Qiagen P100 or P500 kits (according to the manufacturer's instructions). This yielded large quantities of pure DNA in a very short time. DNA prepared by the Qiagen method for sequencing was PEG precipitated before use and DNA for *in vitro* transcription and translations was phenol extracted several times and ethanol precipitated (to remove RNases present during the preparation).

Bacteria were inoculated into 400 ml of selective media and grown overnight. Cells were pelleted (6,000 rpm, 5 min, 4°C), resuspended in 3.5 ml of ice cold sucrose solution (250 g/l sucrose, 50 mM Tris-HCl pH 8), 0.6 ml of freshly prepared lysozyme (10 g/l) was added and incubated on ice for 15 min. Ice cold EDTA (0.25 M, 1 ml) was added, mixed, the tubes held on ice for a further 15 min and 4.5 ml of Triton X-100 solution (2% v/v Triton X-100, 50 mM Tris-HCl pH 8, 62.5 mM EDTA pH 8) was added. The tubes were mixed thoroughly by inversion and the lysate centrifuged immediately (19,000 rpm, 30 min, room temperature). If no pellet was formed at this stage, the lysate was boiled for 10 min, re-centrifuged and the supernatant removed. Ethidium bromide solution (0.5 ml, 10 g/l stock) was added, followed by CsCl (1 g/ml supernatant) ensuring it completely dissolved. This solution was loaded into 8 ml crimp-seal tubes (Du Pont) ensuring that any air was displaced using liquid paraffin and tubes centrifuged in a fixed angle TFT65.13 rotor (40-50,000 rpm, 40 h, room temperature) in a Sorvall (Du Pont) OTD65B ultracentrifuge. DNA was visualised by UV illumination and the plasmid DNA was gently removed using a wide bore syringe needle. The DNA was extracted several times with an equal volume of isopropanol (equilibrated with CsCl-saturated TE) until all the ethidium bromide had been removed. The CsCl was subsequently removed by dialysing the sample against TE, overnight at 4°C. Dialysis tubing had previously been boiled in the presence of EDTA, stored at 4°C in 50% v/v ethanol and was rinsed in water immediately before use. The DNA concentration of the dialysed solution was then measured.

2.3.10 Plasmid Preparation from *K. lactis*.

Nuclear cloning and expression vectors which had been used to transform *K. lactis* were recovered (and subsequently analysed) by transforming competent *E. coli* with extracts made from yeast cells. The plasmids were then prepared as outlined in section 2.3.7 or 2.3.8. The method used to 'rescue' these plasmid was that of Hoffman and Winston (1987). Yeast culture (5 ml) was grown overnight in selective media and the cells harvested (3000 rpm, 3 min). The pellet was resuspended in 400 µl of yeast breaking buffer (2% v/v Triton X-100, 10 g/l SDS, 1 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA), an equal volume of phenol:chloroform:isoamyl alcohol mixture (24:24:1) was added, followed by about 0.3 mg

of acid washed glass beads. The mixture was then vortexed for 2 min, centrifuged (12,000 rpm, 5 min) and 10 µl of the supernatant was then used to transform competent *E. coli* cells

2.3.11 Storage of *E. coli* Strains.

Single colonies of bacteria were grown in 20 ml of selective media. The cells were pelleted (5,000 rpm, 5 min, 4°C), washed in 10 ml of M9 minimal medium and resuspended in 3 ml of M9 medium. Glycerol (sterile) was added to a final concentration of 40% v/v, mixed and 200 µl samples were aliquoted into prechilled Eppendorf tubes and snap-frozen in liquid nitrogen or dry ice-IMS. Cells were stored at -70°C, where they are viable indefinitely.

2.3.12 Storage of *K. lactis* Strains.

K. lactis was grown overnight in 5 ml of minimal medium, an equal volume of preservation mix (12.6 g/l K_2HPO_4, 3.6 g/l KH_2PO_4, 0.9 g/l trisodium citrate, 0.18 g/l $MgSO_4.7H_2O$, 1.8 g/l $(NH_4)_2SO_4$, 20% v/v glycerol) was added and the cells were snap frozen in dry ice/IMS. Samples were then stored at -70°C (indefinitely).

2.4 MANIPULATIONS WITH DNA.

Many of the basic methods used were those outlined originally by Maniatis *et al.* (1982). However, since the publication of this manual many of the techniques have been modified and updated. So for the purpose of clarification the methods used during this study are outlined below.

2.4.1 Polymerase Chain Reaction.

Fragments of DNA which were not amenable to direct cloning were amplified using PCR with primers containing unique restriction enzyme sites to facilitate later cloning. PCR was performed initially using the Perkin Elmer-Cetus DNA Thermal Cycler according to the manufacturers operating instructions, but later a Techne PHC-3 Programmable Heating Block. Initially AmpliTaq *Taq* DNA polymerase (Perkin Elmer-Cetus) was used (with a reaction buffer of 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01 g/l gelatine) but later Promega *Taq* DNA polymerase and NEB Vent Polymerase were used. dNTPs (Pharmacia, Ultrapure, 5 mM stock stored at -20°C) were used at a final concentration (for each dNTP) of 200 µM. The final amount of target DNA present in each reaction was less

than 10 ng. All oligonucleotide primers were synthesised by J. Keyte and D. Langton (Department of Biochemistry, University of Leicester) using an Applied Biosystems 380B Synthesiser and were provided as solutions in aqueous ammonia. Before use they were ethanol precipitated, resuspended in water and the DNA concentration estimated. Primers were used in equimolar concentrations at final concentrations of 0.2-1 μM.

PCR reactions were carried according to the manufacturers instructions and methods outlined by Ehrlich (1983). A typical reaction profile consisted of 25-30 cycles of a denaturation step (95°C, 1 min), an annealing incubation (40-55°C, 2 min) and an extension incubation (72°C, 3 min). All reactions were carried out in a total volume of 50 or 100 μl and were overlaid with liquid paraffin to prevent evaporation. Tubes were heated to 95°C for 5 min to destroy any DNases, and cooled on ice prior to the addition of polymerase. PCR results were analysed by horizontal TAE (see section 2.4.11) agarose slab gel electrophoresis.

2.4.2 Restriction Endonuclease Digestions.

All enzymes used were supplied, in glycerol, by Gibco-BRL, NEB or Pharmacia with their appropriate reaction buffers and were stored at -20°C. The volume of enzyme used in each digestion depended upon the concentration of DNA to be cut and the length of incubation (up to a maximum level of 10% v/v above which the glycerol concentration becomes inhibitory). Digestions were carried out at 37°C (unless otherwise specified) for the appropriate length of time and the reaction halted by the addition of loading buffer (2.5 g/l bromophenol blue, 2.5 g/l xylene cyanol, 30% v/v glycerol, 1 mM EDTA) to 20% v/v and immediate loading onto a gel or freezing, or by phenol extraction.

2.4.3 Preparation of DNase-free RNase.

The removal of contaminating DNases from RNase A preparations was accomplished by boiling (105°C in a PEG bath, 10 min) a 10 g/l solution of RNase, then rapidly cooling on ice. The RNase was then aliquoted and stored at -20°C.

2.4.4 Digestion of RNA in DNA Preparations.

The large amounts of RNA present in certain DNA preparations was removed by the addition of RNase solution (to about 1 mg/l) and incubation at 37°C for the required length of time. This was performed either during DNA preparation (in which case RNase treatment was followed by phenol extraction and ethanol precipitation of the DNA), or after electrophoresis of the DNA. In the latter case, the agarose gel was gently shaken at 37°C, in running buffer containing 1 mg/l RNase.

2.4.5 Filling-in Recessed 3' DNA Ends.

The recessed 3' (sticky) ends of DNA produced by certain restriction enzyme digestion were filled in, yielding blunt-ended DNA fragments using the Klenow fragment of *E. coli* Dpol I. The reaction was performed in a buffer consisting of 50 mM Tris-HCl pH 7.2, 10 mM $MgSO_4$, 0.1 mM DTT, 0.5 g/l BSA (made and stored at -20°C as a 10x stock). All four dNTPs were added to a final concentration of 10 μM and 1-2 μg of DNA was used per reaction. Incubation was carried out at 30°C for 30 min, the solution was phenol extracted, ethanol precipitated and redissolved in sterile TE buffer.

2.4.6 Phenol Extraction.

Phenol extraction was used to remove contaminating protein from DNA preparations or to destroy enzymes used during DNA modifications. Ultra-pure and pre-equilibrated (with Tris pH 7.2) phenol was used as a 24:24:1 mixture with chloroform:isoamyl alcohol and was stored at 4°C. One volume of the phenol phase was added to two volumes of the DNA solution. The two phases were mixed, incubated briefly at room temperature and the tubes centrifuged (11,000 rpm, 3 min, room temperature). The upper aqueous phase containing the DNA was then ethanol precipitated.

2.4.7 Isobutanol Concentration.

Large volumes of dilute DNA solutions were concentrated, prior to ethanol precipitation, by the addition of an equal volume of isobutanol. The two phases were thoroughly mixed, the phases allowed to re-equilibrate and the upper isobutanol phase was then discarded. Extraction was repeated until the aqueous layer attained the required volume.

2.4.8 Ethanol Precipitation of DNA.

Na acetate (3M, pH 5.5) was added to the DNA solution, to a final concentration of 10% v/v, mixed well and two volumes of ice cold absolute ethanol were added. The tubes contents were mixed and placed in a dry ice-IMS bath for 10 min. The DNA was pelleted (11,000 rpm, 10 min, 4°C), the supernatant discarded and the pellet washed briefly with ice cold 70% v/v ethanol. The DNA was dried under reduced pressure for 10 min and resuspended in an appropriate volume of TE buffer (10 mM Tris-HCl pH 7.6, 1 mM EDTA pH 8).

2.4.9 PEG Precipitation of DNA.

Plasmid DNA for sequencing, prepared using the Qiagen kit, was PEG precipitated prior to use. NaCl (4M) was added to the DNA solution to a final concentration of 20% v/v followed by an equal volume of PEG_{8000} (130 g/l). The tubes were incubated (4°C, 30 min), centrifuged (12,000 rpm, 4°C, 30 min), washed in 70% (v/v) ethanol and resuspended in water.

2.4.10 DNA Ligations.

T4 DNA ligase (BRL, 1 U/µl) was used for ligations involving cohesive DNA termini; T4 DNA ligase (Pharmacia, 8 U/µl) was used for blunt end ligations or where very little DNA was available. Reactions were typically carried out in 20 µl volumes containing DNA, 4 µl of T4 DNA ligation buffer (BRL, supplied as 5 x stock) and 1 µl of ligase. Reactions were incubated for 4 h at room temperature or overnight at 15°C, and the complete 20 µl used to transform competent *E. coli*. The volumes of DNA fragments used to obtain the optimal ratio of vector:insert (about 1:2) were calculated taking into account both the concentration and the relative lengths of the two DNA fragments. Transformation, cutting and ligation controls were performed with each experiment.

2.4.11 Electrophoresis of DNA.

DNA fragments were generally analysed by agarose horizontal slab gel electrophoresis, the concentration of agarose varying from 8-15 g/l depending upon the size of the DNA fragment. DNA samples were mixed with an appropriate volume of loading buffer immediately prior to loading onto the gel. Electrophoresis was performed at room

temperature using TAE running buffer (40 mM Tris base, 27 mM Na acetate, 1 mM EDTA, pH adjusted to 8.2 with acetic acid) containing 50 μg/l of ethidium bromide, at 20-90 V, depending upon the agarose concentration and the time span required for resolution of the DNA fragments. DNA in the gel was visualised by long wave UV illumination on a transilluminator. Small DNA fragments (less than 500 bp) were analysed using vertical TBE-PAGE. Gels were made up in TBE buffer (89 mM Tris base, 89 mM boric acid, 2.5 mM EDTA) containing 100-150 g/l acrylamide, 2.5 g/l ammonium persulphate (made fresh as a 100 g/l stock) and 0.0625% v/v TEMED. The gels were stained in ethidium bromide (final concentration 50 μg/l) after electrophoresis.

2.4.12 Elution of DNA from Gels.

DNA was eluted from agarose gels by adsorption to glass milk, using the Geneclean Kit (Bio 101 Inc., California) according to the manufacturer's instructions. The gel slice was melted in NaI (6 M) at 65°C, the DNA adsorbed onto silica glass beads at 4°C and washed and subsequently dissociated from the glass in TE buffer at 65°C. Small DNA fragments (less than 500 bp) were also eluted from agarose or acrylamide gels by electroelution, using an IBI Model UEA Unidirectional Electroeluter. DNA was eluted into 75 μl of salt cushion (7.5 M NH_4 acetate containing 1 g/l bromophenol blue). The elution buffer was 5 mM Tris base, 2.5 mM acetic acid. The salt solution (300 μl) was removed from each well after elution, the DNA was ethanol precipitated and quantified.

2.4.13 Estimation of DNA Concentrations.

The concentration of DNA in a solution was calculated from its OD_{260nm} using a 1 ml (1 cm path length) quartz cuvette in a Shimadzu spectrophotometer. An OD_{260nm} of 1 was assumed to be equivalent to 50 mg/l of ds DNA or 20 mg/l of oligonucleotides.

2.4.14 DNA Size Markers.

The size of DNA fragments produced by restriction enzyme digestion were estimated by the simultaneous electrophoresis of standard DNA fragments. These were either *Hind* III digested (23.13, 9.416, 6.682, 4.361, 2.322, 2.027 & 0.564 kb) or *Eco*R I-*Hin*d III double digested λ genome (21.7, 5.24, 5.05, 4.21, 3.41, 1.98, 1.9, 1.57, 1.32, 0.93, 0.84, 0.58 & 0.14 kb) or *Hae* III digested pUC18 (587, 458, 434, 298, 267, 257, 174, 102, 80, 18 &

11 bp). Concentrated stock solutions were stored at -20°C, and 10-15 µl of these were loaded per lane.

2.4.15 Gel Photography.

Nucleic acid in ethidium bromide stained agarose or acrylamide gels was visualised on a trans-illuminator emitting long wave ultra-violet light and photographed using a Polaroid CU-5 Land Camera with Polaroid Type 667 (black and white) film. Coomassie stained polyacrylamide gels were photographed on a light box using the same camera and film. Later a Mitsubishi Video Copy Processor (model K65HM) was used.

2.4.16a. Oligonucleotide Purification for Gene Resynthesis.

Purification of long oligonucleotides is essential during gene resynthesis, as although the coupling efficiency for each nucleotide is high, the overall efficiency of the synthesis declines rapidly with each subsequent addition. Thus, after 100 or so cycles (ie. for an oligonucleotide 100 bases long), the product consists of a mixture of the full length and slightly shorter oligonucleotides.

The oligonucleotides were ethanol precipitated, resuspended in water and an equal volume of deionised formamide was added. The samples were denatured at 65°C and electrophoresed on a sequencing gel (containing 120 g/l acrylamide). Marker dye solution (DNA loading buffer, section 2.4.11) was loaded in the two end lanes and electrophoresis was halted when the xylene cyanol dye front reached the bottom of the gel. The full length oligonucleotides were visualised by UV shadowing, using a fluorescent thin layer chromatographic plate. The single distinct band was excised from the gel and the oligonucleotides eluted by cutting the gel slice into small pieces and incubating in 5 ml of elution buffer (1 g/l SDS, 0.5 M NH_4 acetate, 10 mM Mg acetate) overnight at 37°C, in a rotary shaker. The eluted oligonucleotides were then isobutanol concentrated, ethanol precipitated and quantified.

2.4.16b Construction Of The Synthetic Gene.

The two DNA strands were then annealed, by mixing equimolar quantities, heating in a PEG bath (105°C) for 5 min, and allowing to cool slowly to room temperature. The annealed oligonucleotides were then used as a template in a PCR type fill-in reaction. The conditions

used were identical to those for normal PCR, although a low annealing temperature (40°C) was used to allow annealing of the small (17 bp) complementary sequences, Vent polymerase was used and only a short programme (5 cycles) was used. The product of the reactions was then phenol extracted, ethanol precipitated, digested with *Eco*R I and *Bam*H I and re-isolated by PAGE.

2.4.17 Analysis of K-plasmid Content.

An overnight yeast culture (5-10 ml), grown in minimal media, was harvested and washed in 1 ml of sterile water. The cells were then resuspended in 500 ml of SED (1.2 M sorbitol, 20 mM EDTA, 50 mM DTT), incubated (30°C, 30 min) and centrifuged (6,000 rpm, 2 min). The pellet was resuspended in 50 ml of $T_{50}E$ (0.5 M Tris-HCl pH 8, 0.5 M EDTA), 20 ml of yeast lytic enzyme (10 g/l) was added and incubated (30°C, 60 min). SDS (10 ml of 100 g/l stock) was added, followed by 20 ml of 5 g/l proteinase K and this was incubated for a further 60 min at 30°C. The tubes were centrifuged (13,000 rpm, 10 min), 20 ml of the supernatant was taken, 5 ml of loading buffer was added and the sample run on a TAE agarose gel (6 g/l).

2.4.18 Preparation of Salmon Sperm DNA.

Salmon sperm DNA (Sigma, sodium salt) was used as a carrier to increase the efficiency of yeast transformation. It was made as a 10 g/l solution in water, allowing 2-4 h to dissolve at room temperature. The solution was then repeatedly forced through an 18 gauge syringe needle, to shear the DNA. The solution was then stored at -20°C in 1 ml aliquots, and denatured immediately prior to use by heating (95°C, 10 min) and rapid cooling on ice.

2.4.19 Southern Analysis.

DNA was transferred from agarose gels to nylon membranes by the method of Southern (1975). DNA was electrophoresed and photographed as usual, but the gel was then soaked sequentially in the following three solutions for 30 min (with brief washes in water after each treatment): Depurinating solution (0.25 M HCl); Southern denaturing solution (0.5 M NaOH, 1.5 M NaCl); Southern neutralising solution (0.5 M Tris-HCl pH 7.4, 1.5 M NaCl). Several pieces of Whatman 3MM filter paper, soaked in 10x SSC (1.5 M NaCl, 0.5 M Na citrate), were placed on top of a number of absorbent sponges lying in a reservoir of 10x

SSC. The gel was carefully placed, upside down (as DNA migrates at the bottom of the gel), on top of the filter paper and its orientation noted. A piece of nylon membrane (Hybond N, Amersham), cut to the exact size of the gel and also pre-soaked in 10x SSC was then placed onto the gel, ensuring no air bubbles were trapped, and its orientation on the gel was noted. Several more pieces of pre-soaked 3MM filter paper, all cut slightly smaller than the nylon membrane, were placed on top of the membrane, followed by a stack of dry absorbent paper towels. Finally, a weight was placed on top of the paper towels to ensure good contact between the gel, membrane and the absorbent tissues. The apparatus was left overnight, at room temperature, to allow the DNA to transfer to the nylon membrane by capillary action. After blotting, the nylon membrane was removed and the efficiency of DNA transfer was assessed by restaining the gel. The DNA on the membrane was immobilised by exposure to UV on a transilluminator for 5 min. The membrane was then placed in a heat sealed hybridisation bag and prehybridised for 6-24 h in 20 ml of Church buffer (10 g/l crystalline grade BSA, 1 mM EDTA, 0.5 M NaH_2PO_4 , pH 7.4, 70 g/l SDS) in a shaking water bath. The prehybridisation and hybridisation temperatures were calculated according to Maniatis *et al.* (1982) or by tempeature = ([2xA.T]+[4xG.C])-20. The fractionated probe was added directly to the pre-hybridisation mix, the bag resealed and reincubated for 18 h. Unhybridised probe was removed from the membrane with three washes in 3 x SSC, 1 g/l SDS (45-65°C, depending upon the melting temperature of the DNA sequence, 30 min each). The membrane was sealed in clingfilm and the hybridised probe visualised by autoradiography.

2.4.20 Colony Hybridisations.

When the efficiency of insertion of a DNA fragment into a vector during plasmid construction was low, large numbers of colonies on agar plates (the products of the bacterial transformation) could be screened for the presence of the required DNA. The method used was adapted from Grunstein and Hogness (1975). Replicas of the colonies were made by momentarily overlaying the plates with a nylon disc (Hybond N, Amersham) and incubating the membranes on fresh plates until the colonies were of sufficient size. The orientation of the filters on the plates were noted, the filters removed and placed sequentially (colony side up) onto Whatman 3MM filter papers soaked in the following four solutions for 5 min each:

0.5 M NaOH; Southern denaturing solution; Southern neutralising solution; 0.5 M Tris-HCl pH 8, 2 x SSC. Filters were firmly blotted dry, the DNA immobilised by exposure to UV on a transilluminator, then prehybridised and hybridised as outlined in Section 2.4.19.

2.4.21 Radiolabelled Probe Preparation.

DNA probes were radiolabelled using random hexanucleotide priming method (Feinberg & Vogelstein, 1983). Reactions were performed in a volume of 15 µl, containing 8 µl DNA fragment, 3 µl oligonucleotide labelling buffer, 1 µl BSA, 1 µl (6 U) Klenow and 2 µl [α-^{32}P] dCTP. Oligo labelling buffer was a mixture of solutions A:B:C in the ratio 10:25:15. Solution A (1 ml): 18 µl β-mercaptoethanol, 5 µl of each 100 mM dATP, dGTP and dTTP; Solution B: 2 M HEPES - NaOH, pH 6.6; Solution C: hexanucleotides (Pharmacia, in TE buffer). The reaction was incubated for 6-20 h at room temperature. The unincorporated nucleotides were then removed by passing the reaction mixture through a Sephadex matrix. A Sephadex G50 (preswollen with TE) column was constructed in a plugged glass Pasteur pipette and equilibrated with 100 µl aliquots of TE. The crude radiolabelled probe mix (volume adjusted to 100 ml with TE) was applied to the column and eluted with 100 µl aliquots of TE. Fractions of 100 µl were collected from the bottom of the column in Eppendorf tubes and the passage of probe through the column was monitored with a Geiger counter. A rise in the radioactivity levels was observed after elution of about 500 µl and the following five fractions with the highest counts were pooled and added to the prehybridisation. Unincorporated nucleotides were retained longer on the column and eluted as a secondary peak.

2.4.22 DNA Sequencing.

DNA was sequenced by the dideoxynucleotide chain termination technique (Sanger *et al.*, 1977) using the T7 Sequencing Kit (Pharmacia) according to the manufacturer's instructions. DNA fragments were labelled with [α-^{35}S] dATP. Sequencing reaction products were analysed on a 0.5 mm x 45 cm x 20 cm (50 ml) acrylamide gel. Commercially available sequencing grade 400 g/l acrylamide (Accugel 40, 19:1 acrylamide: bisacrylamide) was used for making the gel. The urea stock (460 g/l) was deionised prior to use, and was stored, for upto two weeks at 4^{o}C. Sequencing gels (1 x TBE, 230 g/l urea, 70 g/l acrylamide, 1 g/l

ammonium persulphate, 0.1% v/v TEMED) were electrophoresed in TBE buffer (see section 2.4.11) at a constant 1000-1400 V for 2-4 h.

2.4.23 Autoradiography.

When using [α-^{35}S]-dATP in DNA sequencing or [^{35}S]-methionine in protein analysis, gels were dried prior to autoradiography, and exposed naked to X-ray film (Kodak XAR5 13x18 cm, Fuji NIF RX 18x24 cm or Amersham Hyperfilm) at room temperature. With ^{32}P, filters were wrapped in clingfilm and exposed to X-ray film between two intensifying screens at -70^{o}C. Film was exposed overnight, and developed using an Agfa-Gevaert Gevamatic automatic film processor.

2.5 MANIPULATIONS WITH RNA.

All water, tubes and tips used in RNA manipulations were pre-treated with DEPC (0.1% v/v diethyl pyrocarbonate, left at room temperature for 1 h, then autoclaved) prior to use. All electrophoresis equipment used was also kept specifically for RNA work only, to avoid contamination with RNase and was washed with SDS (100 g/l) and rinsed prior to use.

2.5.1 Total Yeast RNA Preparation.

The rapid RNA procedure reported by Schmitt *et al.* (1990) was used to prepare total cellular RNA from yeast. An overnight yeast culture (20 ml) was harvested (5,000 rpm, 5 min) and resuspended in 2 ml of AE buffer (50 mM Na acetate pH 5.3, 10 mM EDTA). A 400 µl aliquot was transferred to a microfuge tube and 80µl of SDS (100 g/l) was added and vortexed for 30 s. An equal volume of phenol (equilibrated with AE buffer) was added, revortexed, the tube incubated (65^{o}C, 4 min) and then immediately chilled on dry-ice IMS until phenol crystals appeared (2-3 min). The heating-cooling cycle was repeated twice, the tubes centrifuged (13,000 rpm, 2 min), and the aqueous phase phenol extracted twice more (at room temperature). The aqueous phase was then ethanol precipitated for 1 h, the RNA resuspended in 50 µl sterile water and 10 µl was loaded per lane, for Northern analysis.

2.5.2 *In vitro* Transcription, using T7 RNA Polymerase.

A typical *in vitro* transcription assay of 20 µl volume was performed in TMS buffer (40 mM Tris-HCl pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine) supplemented with DTT (final

concentration 5 mM), 100 ng of target plasmid, rNTPs (final concentration 3.2 mM) and RNA guard (1 µl of Pharmacia placental RNase inhibitor, 20 U/µl). T7 Rpol was initially provided by Mr. P. Reay (Leicester Biocentre) and was used at 1 µl/ 20 µl reaction (specific activity unknown); later Pharmacia FPLC purified T7 Rpol (70 U/µl) was used at 1 µl/ 20 µl reaction. Reactions were incubated (37°C, 30-60 min) and the reaction stopped by the addition of sterile loading buffer (50% v/v glycerol, 1 mM EDTA, 4 g/l bromophenol blue, 4 g/l xylene cyanol) to 20% v/v.

2.5.3 RNA Agarose Gel Electrophoresis.

RNA was analysed by horizontal slab gel electrophoresis on a denaturing agarose gel, consisting of 15 g/l agarose, 2.85% v/v formaldehyde in MOPS buffer (3 mM Na acetate, 1 mM EDTA, 20 mM MOPS, pH 7). Formaldehyde was added to the agarose just prior to pouring, to minimise loss by evaporation; for the same reason gels were only immersed in MOPS running buffer immediately prior to electrophoresis. Gels were electrophoresed at 50-100 V for 2-5 h at 4°C.

2.5.4 RNA Sample Preparation.

Freshly made sample cocktail 15 µl (0.5 µl 20x MOPS, 10 µl deionised formamide, 3.5 µl of 38% v/v formaldehyde, 0.1 µl 10 g/l ethidium bromide) was added to 5 µl of RNA sample, any ds RNA denatured (by heating to 65°C, 10 min, and immediately cooling on ice) and 5 µl of sterile loading buffer was added.

2.5.5 RNA Size Markers.

The size of transcripts was estimated by the simultaneous electrophoresis of 5 µl of standard RNA fragments (0.24-9.4 kb RNA ladder, Gibco-BRL). This was prepared for use as outlined for RNA samples, above.

2.5.6 Northern Blotting.

RNA was electrophoresed and transferred to a nylon membrane for analysis, using a modified version of the method of Bostian *et al.*, (1983). This was almost identical to that used for Southern blotting except that the RNA gel required no pre-treatment after electrophoresis and prior to blotting. The RNA agarose gel was also placed upside down, as

RNA, like DNA, migrates at the bottom of the gel and the RNA was transferred to the membrane by capillary action. The efficiency of the transfer of RNA to the membrane was determined by examination of the gel and the membrane under UV illumination. RNA was immobilised on the membrane by exposure of the appropriate membrane face to UV light on a transilluminator for 4 min, and the membrane pre-hybridised and hybridised as outlined in Section 2.4.19

2.6 MANIPULATIONS WITH PROTEINS.

2.6.1 Analysis of Heterologous Protein Production in *E. coli*.

2.6.1.1 *Staphylococcal* Protein A Fusion Vectors.

The plasmid vectors pRIT2T and pRIT5 were designed to allow expression of heterologous proteins in *E. coli* as carboxy-terminal fusions with *Staphylococcal* protein A (Nilsson *et al.*, 1985). This also facilitated subsequent purification, the protein A moiety binding to an IgG affinity matrix.

2.6.1.1a Sample Preparation.

E. coli was grown overnight at 30°C in 100 ml of selective media, sub-cultured (1% v/v) into fresh media and allowed to continue growing at 30°C until mid log phase (OD_{600nm} approximately 0.6, 3-4 h). The culture was then split into two portions, one of which was induced by a temperature shift to 42°C (a rapid shift was attained by the addition of an equal volume of media at 54°C) and allowed to incubate for a further 2 h at 42°C; the other was kept at 30°C as an uninduced control. Cells (5 ml) were pelleted (6,000 rpm, 5 min, 4°C), resuspended in 100 µl of ice cold water and the protein content analysed by SDS-PAGE. The remainder of the culture was used for affinity chromatography. The cells were pelleted (as above), resuspended in ice cold TST (5 ml) and lysed by sonication on ice (four 30 s bursts, with 30 s intervals on ice) or by four cycles of freeze-thaw (dry ice-IMS) and vortexing with glass beads. Unlysed cells and cellular debris were removed by centrifugation (18, 000 rpm, 10 min, 4°C).

2.6.1.1b IgG Sepharose Affinity Chromatography.

The IgG Sepharose matrix (Fast Flow, Pharmacia) was packed into a 2 ml column and washed with 10 ml of TST (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.05% v/v Tween

20). Before use the column was equilibrated by the sequential addition of 0.5 M acetic acid pH 3.4 (6 ml), TST (6 ml), 0.5 M acetic acid (6 ml) and TST (6 ml), at a flow rate of about 30 ml/h. Samples, derived from 50-100 ml of cultures, were applied when the pH of the eluate attained neutrality and were washed through the column with TST. The passage of proteins through the column was monitored by a UV detector linked to a chart recorder. Loosely bound proteins were eluted by the addition of 5 mM NH_4 acetate pH 5 (4 ml) at a reduced flow rate of about 15 ml/h. More specifically bound proteins were then eluted with 0.5 M acetic acid pH 3.4 and 1 ml fractions collected. The protein content of each fraction was quantified (Section 2.6.2b) and analysed by SDS-PAGE. The column was recycled by washing with TST and stored at 4°C.

2.6.1.1c Insoluble Protein Analysis.

Analyses of the more insoluble protein components was performed using the following method (Brophy, Pers.Comm.). Cultures (50 ml) were grown and induced (Section 2.6.1.1a). Cells were pelleted (6,000 rpm, 5 min, 4°C), resuspended in 3 ml of lysis buffer (50 mM Tris-HCl pH 8, 1 mM EDTA, 100 mM NaCl, 0.5 mM PMSF) containing lysozyme (final concentration 5 g/l) and incubated (5 min, room temperature). The cell suspension was then sonicated for four 30 s bursts on ice (with 30 s intervals). The unlysed cells and cell debris were removed by gentle centrifugation (1500 rpm, 3 min, 4°C) to ensure any inclusion bodies formed were not sedimented. The pellet was retained for analysis (sample A). The supernatant was then removed, centrifuged (11,000 rpm, 10 min, 4°C) and this supernatant retained for analysis (sample B). The pellet was resuspended in 1.5 ml of fresh 2 M urea (in lysis buffer) and incubated (room temperature, 5 min), recentrifuged (as above) and the supernatant retained (sample C). The pellet was then resuspended in 1 ml of 0.5% v/v Triton X100, 10 mM EDTA (in lysis buffer). This was centrifuged after incubation (5 min, room temperature) and the supernatant retained for analysis (sample D). The pellet was washed in 1 ml of ice cold water and resuspended in 100 µl of cold water (sample E). Samples of all supernatants were then analysed by SDS-PAGE.

2.6.1.2 *Schistosomal* GST Fusion Vectors.

The pGEX plasmids were constructed to allow one step purification of heterologous proteins expressed as carboxy-terminal fusions with *Schistosomal* glutathione-S-transferase (Smith and Johnston, 1987; Smith and Corcoran, 1990). Purification was accomplished by the use of a glutathione affinity matrix.

2.6.1.2a GST Affinity Matrix Preparation.

The glutathione agarose affinity matrix (sulphur linked; Sigma) was used either packed into a column or in Eppendorf tubes. The latter was particularly useful for rapid small scale analysis of recombinant protein in cell extracts. For column chromatography the matrix was pre-swollen and packed into a 1 x 10 ml bed volume vertical column and once packed was stored at 4°C in Na azide (0.2 g/l). Immediately prior to use the column was pre-equilibrated with 100 ml of MTPBS (150 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4, pH 7.3) or solublizing buffer (50 mM Tris-HCl pH 8, 50 mM DTT, 10 mM EDTA, 100 mM NaCl, 1% v/v Triton X100, 0.3 g/l SDS) at a flow rate of 10-15 ml/ h. The column was recycled by washing with 30 ml of NaCl (3 M), followed by 50 ml of Na azide (0.2 g/l).

2.6.1.2b Sample Preparation for Glutathione Affinity Chromatography.

E. coli were grown overnight in selective medium, sub-cultured (1% v/v) into 20 ml of fresh medium and incubated until the culture attained an OD_{600nm} of about 0.6. Cell suspensions were then divided into two portions, one of which was induced by the addition of IPTG to a final concentration of 0.4 mM and incubated for a further 1-18 h. Cells (2 ml) were pelleted (6,000 rpm, 5 min, 4°C), resuspended in 50 µl of ice cold water and analysed by SDS-PAGE. The remainder of culture was pelleted and resuspended in 0.5 ml of ice cold MTPBS. The cells were lysed on ice, as above (section 2.6.1.1a). Triton X100 was added (final concentration 1% v/v), the tubes centrifuged (11,000 rpm, 5 min, 4°C) and the supernatant then applied to the pre-equilibrated GST affinity column. Samples were washed through with MTPBS. Adsorbed proteins were eluted by the addition of 20 ml of freshly prepared reduced glutathione (10 mM, in 50 mM Tris-HCl pH 8) and 1 ml fractions were

collected. Each fraction was then analysed for protein concentration (Section 2.6.2b) and by SDS-PAGE (Section 2.6.2a).

The more insoluble proteins were analysed by resuspending the cell pellet was in ice cold solublisation buffer (3 ml, Section 2.6.1.2a). The cells were then lysed and treated as above (section 2.6.1.1a). Samples were then applied and eluted using solublisation buffer.

2.6.1.2c Rapid Mini-preparation of GST Fusion Proteins.

E. coli were cultured and induced as outlined above. Cells were then centrifuged (5,000 rpm, 5 min, 4°C) and resuspended in MTPBS (4 ml). The suspension was sonicated for two 30 s bursts on ice and centrifuged (12,000 rpm, 5 min, 4°C). The supernatant was removed, 100 µl of the affinity matrix (in Na azide) was added and incubated on ice for 30 min. The matrix was then pelleted (6,000 rpm, 2 min, 4°C), the supernatant discarded and the pellet washed three times in 1 ml of ice cold MTPBS. Finally the matrix was resuspended in 100 µl of ice cold water and analysed by SDS-PAGE. This was boiled (105°C, 10 min), centrifuged (12,000 rpm, 2 min, room temperature) and the samples analysed by SDS-PAGE.

The protein components which were not amenable to affinity chromatography were analysed using the method outlined in Section 2.6.1.1c, except that protein synthesis was induced with 0.4 mM IPTG (final concentration) rather than heat shock.

2.6.1.2d Identification of the protein isolated by glutathione affinity chromatography.

GST catalyses the reaction between glutathione and 1-chloro-2,4-dinitrobenzene (CDNB) yielding a coloured product. The course of the reaction can, therefore, be followed by optical density measurements at 340 nm. All the components were added (20µl 50 mM glutathione, 20µl of 50 mM CDNB, 500 µl MTPBS, 100 µl sample, 360 µl water) the tubes mixed and incubated at room temperature for 10 min.).

2.6.1.3 T7 RNA Polymerase Expression Vectors.

The pET plasmids (Rosenburg *et al.*, 1987; Studier *et al.*, 1990) were constructed to facilitate the high level expression of heterologous genes when inserted between the powerful expression

signals of bacteriophage T7 gene 10. They were designed such that, when used in conjunction with the appropriate *E. coli* host strain BL21(DE3), containing a chromosomal copy of the IPTG-inducible T7 Rpol gene) and the antibiotic rifampicin (which inhibits cellular Rpols), only genes immediately downstream of T7 promoters are transcribed and therefore, expressed.

2.6.1.3a Sample Preparation.

E. coli were grown and induced as outlined in Section 2.6.1.2b After 2-4 h of induction, cells were harvested (5,000 rpm, 5 min, 4°C), resuspended in 500 µl of solublisation buffer (Section 2.6.1.2a) and lysed by sonication. Cell debris and unlysed cells were pelleted (11,000 rpm, 5 min, 4°C) and the supernatant analysed by SDS-PAGE. Those proteins which were not solublised by this method, were analysed using the method outlined in Section 2.6.1.1c.

2.6.1.3b Radiolabelling of Recombinant Proteins.

E. coli were grown overnight in selective medium, then washed and resuspended in M9 media containing ampicillin and the amino acid supplement (Section 2.1.2). Cell suspensions were then incubated for a further 3-4 h. Half of the culture was then induced with IPTG (final concentration 0.4 mM), the other half remaining uninduced; both were then reincubated (37°C, 30 min). Rifampicin (20 mg/ml stock, in methanol) was added (final concentration 200 µg/ml), the cultures incubated at 42°C for 30 min (rifampicin is more effective at this temperature) and were then returned to 37°C for a further 30-90 min. A 1 ml sample of cells was transferred to an Eppendorf tube and [^{35}S] methionine (10 µCi) was added, and the tubes reincubated (37°C, 10 min). Cells were pelleted (11,000 rpm, 5 min, room temperature), resuspended in 50 µl of water and the protein content analysed by SDS-PAGE.

2.6.1.3c *In vitro* Transcription-Translation.

This was performed using the *E. coli* S30 Extract Prokaryotic Translation System for Linear DNA (Promega), according to the manufacturer's instructions. Only ultra-pure plasmids, prepared by either CsCl gradient sedimentation or Qiagen columns, were used. In the latter

case, the DNA solution was phenol extracted and ethanol precipitated prior to use, to remove the RNase added during plasmid preparation.

2.6.2a SDS-PAGE.

Discontinuous vertical slab gel SDS-PAGE (Hames, 1981) was performed using either a 40 ml BRL apparatus or a 10 ml Hoeffer mini-gel apparatus. The acrylamide concentration of the resolving gel varied from 100-200 g/l according to the size of the proteins analysed. A premixed 300 g/l acrylamide (37.5:1 acrylamide: bisacrylamide) stock was used for 10-15% w/v gels; a 400 g/l stock (ratio 37.5:1) was used for 20% w/v gels. All components (except TEMED) were mixed and degassed under reduced pressure for 5 min immediately prior to pouring the gel. Resolving gels were made, poured and overlaid with water saturated isobutanol until set (30 min - 1 h). A typical resolving gel contained 0.375 M Tris-HCl pH 8.8, 1 g/l SDS, 1 g/l ammonium persulphate, 0.05% v/v TEMED. The concentration of the acrylamide was varied by dilution of the stock solution with water. When the resolving gel had polymerised, the isobutanol was poured off, the top of the gel rinsed with water and the stacking gel (125 mM Tris-HCl pH 6.8, 1 g/l SDS, 45 g/l acrylamide, 1% g/l ammonium persulphate, 0.05% v/v TEMED) poured on top. The comb was then inserted and the gel allowed to set completely (about 1 h).

Samples were prepared by the addition of an equal volume of sample loading buffer (78 mM Tris-HCl pH 6.8, 100 g/l SDS, 12% v/v β-mercaptoethanol, 2 g/l bromophenol blue, 40% v/v glycerol) heated in a PEG bath (105°C, 10 min) and centrifuged (12,000 rpm, 2 min, room temperature) immediately prior to loading. Volumes of 10-50 µl was loaded per well in the BRL apparatus, 5-20 µl in the Hoeffer, depending upon protein concentration. The electrophoresis buffer used was 25 mM Tris base, 192 mM glycine, 1 g/l SDS.

Gels (40 ml) were run at 20-30 V overnight or 80-100 V for 5-7 h (depending upon gel concentration); Hoeffer mini-gels were run at 20 mA for 1-2 h. After electrophoresis, gels were fixed and stained (40% v/v methanol, 10% v/v glacial acetic acid, 1.2 g/l Coomassie Brilliant Blue R) for 60 min and then destained in fixing solution (40% v/v methanol, 10% v/v glacial acetic acid). Gels which required autoradiography were fixed in destaining solution for 30 min, rehydrated in water for 15 min, immersed in rapid autoradiographic

enhancer (Enlightning, Du Pont NEF-974) for 30 min, dried down under vacuum at 80°C for 2-4 h (Biorad Dual Temperature Slab Gel Dryer Model 1125B) onto Whatman 3MM filter paper then autoradiographed.

2.6.2b Calculation of Protein Concentrations.

Protein concentration was estimated according to the method of Bradford (1976) using a BSA or ovalbumin standard with the Biorad Protein Assay, according to the manufacturer's instructions.

2.6.2c Protein Molecular Weight Markers.

Protein molecular weight markers (29-205 kDa, Sigma, SDS-6H) were used, according to the manufacturer's instructions. ^{14}C-labelled Rainbow molecular weight markers (14.3-200 kDa, Amersham) were used for SDS-PAGE autoradiography.

2.6.3 Western Analysis.

Protein to be immunologically analysed was immobilised on a nitrocellulose membrane either by direct application (in the case of synthetic peptides) or from SDS-polyacrylamide gels by 'wet' electroblotting (Harlow & Lane, 1988) using the Hoeffer Baby Blotter. A piece of nitrocellulose (0.2 μm pore size, Fisons) cut to the size of the gel, along with six pieces of Whatman 3MM filter paper of slightly larger dimensions and the Scotchbrite support pads, were pre-soaked in blotting buffer (25 mM Tris base, 192 mM glycine, 25% v/v methanol) for 40 min. After electrophoresis the gel was placed onto the membrane, was sandwiched between the Whatman 3MM and finally the Scotchbrite supports, ensuring no air bubbles were trapped. The gel-membrane sandwich was then assembled in the electroblotting apparatus, ensuring the membrane-gel was correctly oriented The net negative charge of the SDS-protein complex means that samples migrate towards the anode where they are immobilised on the membrane. The apparatus was filled with blotting buffer and blotting was performed at 90V (2-3 h, 4°C) or overnight (30V, 4°C). After blotting, the efficiency of protein transfer was checked by Ponceau S (20 g/l 3-hydroxy-4-[2-sulpho-4-(sulphophenylazo)phenylazo]-2,7-naphthalene disulfonic acid, 300 g/l trichloroacetic acid, 300 g/l sulfosalicylic acid) staining the membrane, for 2 min at room temperature. This also allowed the position of the lanes and the molecular weight markers to be noted. A similar

procedure was used for visualising the synthetic peptides which were applied directly to the membrane. The membrane was destained by several washes in phosphate buffered saline (PBS; 137 mM NaCl, 3 mM KCl, 2 mM NaH_2PO_4, 10 mM Na_2HPO_4, pH 7.2).

All further incubations during immunological 'probing' were carried out in PBS with gentle agitation for 2 h at room temperature, and the filter was thoroughly washed in 200 ml of PBS prior to commencement of each subsequent stage. The nitrocellulose membrane was initially 'blocked' by incubation in 100 ml of Marvel solution (10 g/l, in PBS). The filter was then washed and incubated with the primary antibody (harvested from chicken egg yolk, Section 2.6.6) at the appropriate dilution. Alternatively, the primary antibody was added directly to the blocking solution. The filter was then rewashed and incubated with the secondary antibody (goat anti-chicken antibody, at a dilution of $1x10^{-3}$ v/v) and finally, after washing the filter again, the tertiary antibody (rabbit anti-goat-conjugated to peroxidase, at dilutions of 3 to $5x10^{-4}$) was added and the filter incubated in the dark. The filter was then washed and analysed by the addition of the developing solution (0.75 g/l chloronaphthol, 20% v/v methanol, 0.0005% v/v H_2O_2, in PBS). Development of the filter was halted by washing in ice cold water, and drying at room temperature.

2.6.4 Preparation of Synthetic Peptides.

Predictions concerning the possible antigenic sites on the putative K2 DNA polymerase were made by computer analyses on the basis of a number of different criteria (antigenicity, surface probability, hydrophilicity, and flexibility prediction) by Dr. M. Crossman (The Advanced Biotechnology Centre, Charing Cross and Westminster Medical School, London). The peptides were synthesised on a 4-branch MAP poly-lysine core (Applied Biosystems Inc.) by Mr. T. Seddon and Dr. K. Lilley (Department of Biochemistry, Leicester University) and their structure confirmed by GC-MS.

2.6.5 Immunisation of the Chickens.

Two laying hens were purchased and maintained by Biomedical Services (University of Leicester), who subsequently performed all the necessary injections and collected the eggs. The eggs were stored at 4°C until required. The synthetic peptides were supplied as a powder and were dissolved in PBS. Synthetic Peptide 1 (SP1, QIDDYSDD, 5.7 mg) was

dissolved in 3 ml of PBS, 100 ml of this was taken and diluted to 750 ml. An equal volume of Freunds Complete Adjuvant was added and the entire 1.5 ml (ie. 190 μg of peptide) was injected into the pectoral muscle. Further identical samples were then injected 12 and 20 days later. The same procedure was followed for Synthetic Peptide 2 (SP2, TYGDINKD) but only 150 μg of peptide was administered at each injection.

2.6.6 Chicken Antibody Preparation.

Antibodies preparations, from eggs collected 25-35 days after the original injection, were prepared according to the method of Gassman *et al.*, (1990). Four egg yolks were washed in distilled water and the skins removed. PBS (60 ml) was added, followed by an equal volume of PBS containing 70 g/l PEG_{6000} and was stirred at room temperature for 30 min. This was then centrifuged (14,000 rpm, 10 min, 4°C) and the supernatant poured through a milk filter. Solid PEG_{6000} was added to a final concentration of 120 g/l and was stirred at room temperature until it all dissolved. This was then recentrifuged (as above) and the pellet resuspended in 40 ml of PBS. An equal volume of PBS containing 240 g/l PEG_{6000} was added, mixed thoroughly and centrifuged (as above). The pellet was then resuspended in 20 ml of PBS and dialysed against PBS (overnight, 4°C). The dialysate was recentrifuged (as above) to remove any particulate material and the crude antibody solution was stored at 4°C.

CHAPTER 3

ATTEMPTED OVER-EXPRESSION OF THE PUTATIVE DNA POLYMERASES IN *E. COLI*.

3.1 INTRODUCTION.

In view of their resemblances to other 'linear plasmids', the manner in which the K-plasmids replicate is of particular interest, and dissection of this system would be facilitated by access to the purified replicatory components. The initial strategy for isolating sufficient quantities of these putative proteins was, therefore, based upon the microbial over-expression of the K-plasmid ORFs, using recombinant DNA technology. The production of foreign proteins in these systems, many of which were previously unavailable due to their limited supply from natural sources, has aided research in many disciplines and led to large scale exploitation in some cases.

Most work on the synthesis of heterologous proteins has focused upon either *E. coli* or *S. cerevisiae* as model systems for prokaryotic and eukaryotic gene expression (reviewed by Goeddel, 1990; Romanos *et al.*, 1992). Many of the fundamental processes involved in gene expression appear to be similar, at the molecular level, in both *E. coli* and *S. cerevisiae,* despite the former being a prokaryote and latter, a eukaryote. However, there remain many basic biochemical differences. Although this work has concentrated upon heterologous gene expression in *E. coli*, in those instances where experimental evidence generated with *S. cerevisiae* may reflect the existence of similar mechanisms in *E. coli* (eg. transcription, translation and codon bias) these are discussed. Furthermore, due to the enormous volume of literature concerning gene expression in these systems, and in order to keep this section as brief as possible, much of the information outlined is not specifically referenced. Instead, I have referred to the appropriate review articles. For the same reason, a detailed analysis of the factors contributing to successful foreign gene expression is not in the scope of this article. The following text, therefore, contains just a broad outline of the processes involved.

The over-expression of a non-essential foreign gene imposes a significant metabolic burden on the host cell, and often results in reduced cellular growth rate. In constitutive expression systems, where there is no control over the regulation of heterologous gene expression, cellular growth and gene expression are linked and there is, therefore, a considerable selection for cells exhibiting reduced foreign gene expression. Under these circumstances it

is desirable to use tightly regulated promoters to control gene expression, and a wide range of plasmid vectors have been developed to allow the precise control of such protein synthesis. Despite the apparent widespread success of these techniques, the failure to efficiently over-produce certain proteins is fairly common, but generally goes unreported. Problems can arise at any stage in the complex process of gene expression, and the reasons for these failures remain unclear.

As the number of cellular ribosomes far exceeds the abundance of any one class of mRNA, gene expression can be maximised by increasing the proportion of the target transcript. This can be accomplished by either increasing the plasmid copy number (Section 3.1.1), increasing promoter strength (Section 3.1.2) or increasing the useful lifetime of the transcript (Section 3.1.4).

3.1.1 Gene Dosage and Plasmid Stability.

In general, the higher the number of copies of a heterologous gene, the greater the yield of product. However, duplication of sequences (eg. tandem copies of the foreign gene) within a plasmid are avoided due to complications which arise with regards to plasmid structural stability, often as a result of intramolecular recombination. Many expression vectors are, therefore, based on high copy number cloning plasmids, often pBR322 or pUC (reviewed by Balbas & Bolivar, 1990). However, the high level expression of foreign genes on multicopy plasmid is a great burden to the host cell, so in the absence of strong selective pressure, plasmid-less segregants rapidly arise and outgrow those retaining the plasmid. Such 'instability' is generally minimised by using antibiotic selection.

3.1.2 Transcription.

This, the first stage of gene expression, involves the Rpol-catalysed interpretation of the DNA sequence, and synthesis of a complementary intermediate (messenger) RNA (reviewed by Gralla, 1990). Gene expression tends to be regulated primarily at the transcriptional level in both *E. coli* and *S. cerevisiae*, and is, therefore, often thought to be the rate limiting step for heterologous gene expression; a steady state mRNA level being the primary determinant of the final product yield.

Transcription, consists of three main stages: initiation, elongation and termination, of which initiation is the critical control point. Initiation involves non-specific DNA binding and sequence 'scanning' by the Rpol holoenzyme. When promoters are recognised, the enzyme is held in a loose, but specific, association. The core promoter consists of four main conserved elements, which have been identified by functional and statistical analyses of a large number of *E. coli* promoters: the -10 and -35 boxes (relative to the transcriptional start site at +1), and the spacer regions between the -10 and -35 motifs, and between the -10 box and transcriptional start site. These conservations refer to specific sequences with regards to the -10 and -35 boxes, and to the lengths of the spacer regions. The simultaneous contact of the Rpol with the two core promoter sequences (or regulatory proteins, or both) results in a series of conformational changes within the promoter, leading to the melting of a short stretch of DNA. This exposes the transcriptional start site, allowing the formation of a stable open complex and the synthesis of a small RNA. This is subsequently elongated when the Rpol breaks the contacts with the promoter and escapes downstream. Generally the formation of the stable open complex is the rate limiting factor in transcription, mRNA synthesis usually being too rapid to limit expression. Nevertheless, in some cases the Rpol does pause during elongation, sometimes for upto several seconds and this can, in extreme cases, contribute to reduced gene expression. Strong pauses sites can be identified by analysis of the mRNA distribution along a gene but elimination of such sites maybe difficult, as these sequences often reside within the coding region.

Experiments with both *E. coli* and *S. cerevisiae* genes have illustrated the importance of upstream and downstream DNA sequences upon gene expression. Upstream activator regions are normally binding sites for regulatory proteins which may interact with the Rpol, altering its properties; in the simplest cases by providing additional stabilising interactions. Downstream activator sequences have been shown to affect the rate of transcriptional initiation using *S. cerevisiae* promoters, resulting in the heterologous genes being expressed at much lower levels than the homologous gene. Such sequences may also exist in *E. coli.*

Transcriptional terminators determine the point at which the mRNA-Rpol-DNA complex dissociates, and are essential in expression vectors. These generally consist of a G+C rich

region which assumes a hairpin secondary structure, followed by an A+T rich region. This causes the Rpol to pause or fall off the template, depending upon whether it is a factor-dependent or independent terminator. Transcriptional terminators not only limit energy expenditure and sequestering of cellular factors in unnecessary transcription and translation, but also isolate the replicative functions of the host plasmid from the effects of strong transcription. Transcriptional 'read-through' from strong heterologous promoters can lead to promoter or *ori* occlusion, disrupting the functioning of other plasmid-borne genes. The production of unnecessarily long transcripts, may also allow the formation of undesirable secondary structures that can reduce translational efficiency.

3.1.4 mRNA Stability.

Transcription and translation appear to be almost simultaneous and experimental evidence indicates that mRNA degradation starts almost immediately and proceeds in same direction. Upon induction, the mRNA of a new gene may typically appear in *E. coli* within 2 min, and the corresponding protein within a further minute. The functional half-life of the average *E. coli* transcript is about 2 min and the factors which influence the stability of mRNA include the presence of ribosomes, the rate of translation, mRNA sequences and secondary structure at the 3' and 5' ends, and the organisms growth rate (reviewed by Brawerman, 1987). The stability of transcripts can be increased by using host stains which are defective in transcriptional termination (*rho*$^-$) or exhibit altered RNA metabolism, or by the inclusion of strong rho-independent terminators. Highly structured mRNA termini may simply act as barriers against exonucleolytic degradation.

3.1.5 Translation.

This involves the conversion of the informational content of a mRNA into a specific sequence of amino acids, the building blocks of a protein. Translation, like transcription, can also be regarded as consisting of three broad stages: initiation, elongation and termination (recently reviewed by Gold & Stormo, 1990; Schoner *et al.*, 1990; de Boer & Hui, 1990). Initiation is, again, the critical control point and consists of the ribosome binding to the messenger, allowing translation to commence at the appropriate codon. Initiation therefore

involves the interaction of three different RNA species, mRNA, 16 S rRNA of the 30 S ribosomal sub-unit, and the initiating tRNA.

The binding of a mRNA to a ribosome is one of the major rate limiting steps in translation. The rate of translation can, therefore, be maximised by providing an efficient ribosome binding site on the mRNA, in the context of optimal spacing with the initiating codon. Ribosome binding is mediated by the interaction of complementary nucleotide sequences in the 5' untranslated sequence of the mRNA (Shine-Dalgarno sequences) and the 16 S rRNA. This RNA-RNA interaction and its positioning during ribosome assembly subsequently defines the reading frame. The start codon always specifies methionine or N-formyl methionine, and AUG is the preferential initiating codon although others are sometimes used but at lower efficiencies.

There have been reports that the nature of the RNA immediately upstream and downstream of the initiating codon influences the efficiency of translation, although the basis of this is poorly understood. The region upstream of the Shine-Dalgarno sequence may, for example, simply serve a protective function against 5'-3' exonucleases. The spacer between the initiating codon and the Shine-Dalgarno sequence, may influence the efficiency of initiation by minimising secondary structure in this region, which could limit ribosome accessibility.

Translational termination (peptide release from the ribosome) is mediated by one of two factors, RF1 (which recognises UAA and UAG) or RF2 (which recognises UAA and UGA). These have different efficiencies of termination and there is a strong bias towards the UAA non-sense codon in highly expressed genes.

3.1.6 Product Stability.

Both foreign and native proteins are prone to degradation in *E. coli* and short peptides, produced as a result of non-sense mutations are known to be rapidly degraded. Furthermore, the synthesis of abnormal, prematurely terminated or denatured, proteins often induces the proteolytic enzymes which degrade them. Cellular proteases act like a primitive immune system, eliminating unrecognised proteins, but also provide a source of amino acids and nutrients. Problems with proteolysis *in vivo* can usually be circumvented by

synthesising peptides as fusion proteins (reviewed by Uhlen & Moks, 1990) or by inhibition of cellular proteases (reviewed by Gottesman, 1990).

The cellular localisation of a recombinant protein can also affect its stability; for example proteins which are labile in the cytoplasm may be significantly more stable in the periplasm. Protein secretion may, therefore, circumvent such instability (reviewed by Stader & Silhavy, 1990).

3.1.7 Codon Bias.

The codon preference in unicellular organisms such as *E. coli* and *S. cerevisiae* is a constraint imposed by the nt composition of the genome and the population of iso-accepting tRNAs. Strongly expressed genes show a distinct bias towards a codon subset, tending to use those codons corresponding to the most abundant cognate tRNAs (Ikemura, 1982, Varenne & Lazdunski, 1986, Sharp & Li, 1986; de Boer & Kastelein, 1986). A codon adaption index has been developed to allow semi-quantitative measurement of the bias toward certain codons in a gene (Sharp & Li, 1987), and this may be useful for predicting the level of heterologous gene expression. Despite previous protestations to the contrary, based on observations made with the *E. coli lacZ* gene, the contribution of codon bias to efficient expression of heterologous genes is now becoming increasingly clear (see Section 3.3 and Chapter 4).

3.1.8 Product Solubility.

The over-expression of many recombinant genes in *E. coli* is often limited by the production of insoluble cytoplasmic or periplasmic protein aggregates (reviewed by Schein, 1990). The occurrence of this, however, generally remains unpredictable. These 'inclusion bodies' can often be seen by phase contrast microscopy, and can usually be separated from cell lysates simply by centrifugation, facilitating the purification of such products. The protein can then be solublised using strong chaotrophic reagents, but any subsequent biological activity is dependent upon its correct refolding after removal of the denaturant, a process which is usually difficult if not impossible (reviewed by Kohno *et al.*, 1990).

3.1.9 Host Physiology.

Every expression system consists of at least two components: the vector and the host. The host's genetic background is obviously an important factor in the yield of the product, at the level of both plasmid and product stability. In addition, the hosts growth requirements (O_2, temperature, nutrients, pH) and mode of gene regulation may also be crucial in determining the level of gene expression. Increased levels of gene expression may also reduce growth rate and lead to morphological changes (eg. filamentation). On several occasions, the same expression vector has been observed to give widely different yields of the same product in different *E. coli* host strains. Again, the basis for this is not understood.

3.2 RESULTS.

3.2.1 Over-Expression Of The K-plasmid DNA Polymerases.

During this study the over-expression of the putative K-plasmid Dpol genes, or fragments thereof, has been attempted as fusion and non-fusion products. The pTTQ and pET vectors were used to try to express the native gene; the pRIT2T, pRIT5 and pGEX vectors were used to attempt expression as fusion products. As previously discussed, factors taken into account during the construction of these plasmids include the provision of strong regulated transcriptional and translational initiation and termination sites and features to increase protein and plasmid stability and to facilitate isolation of the final product. The pTTQ and pGEX vectors, and the pET host strain all carry the *lac*I^Q gene and, therefore, give tight repression of heterologous genes under non-inducing conditions. The pTTQ and pGEX plasmids contain the strong synthetic *trp-lac*UV5 (*tac*) promoter. The *lac*UV5 promoter bears a double point mutation in the Pribnow box, increasing promoter strength but also reducing its sensitivity to catabolite repression. In the *tac* promoter, the -35 region of *lac*UV5 promoter has been replaced with that of the *trp* promoter. The pGEX and pRIT vectors provide additional protein moieties to increase product stability and facilitate isolation of soluble products by affinity chromatography. The pET vectors contain strong T7 transcriptional and translational initiation and termination sequences, and allow detailed analysis of product synthesis, both *in vitro* and *in vivo*.

3.2.1 Expression As Non-Fusion Proteins.

3.2.1.1 The pTTQ Vectors.

These were designed and constructed at the Leicester Biocentre (Stark, 1987) specifically to allow inducible high level expression of cloned genes, but also to ensure tight repression of these genes under non-inducing conditions. Expression is driven by the *tac* promoter, fused to the translational and extreme 5' coding sequence of the *E. coli* β-galactosidase gene. The pTTQ vectors were used to express the *luxA* and *luxB* subunits of the *Vibrio harveyi* luciferase gene, the Tn9 chloramphenicol resistance (chloramphenicol acetyltransferase) gene and the Tn903 kanamycin resistance gene (Stark, 1987). Derivatives of these vectors were also used to clone and over-express the genes for the *E. coli* gyrase A and B sub units (Hallet *et al.*, 1990).

3.2.1.1a Cloning Of The K1 And K2 DNA Polymerase Genes.

The high genomic A+T content of the K-plasmids resulted in the presence of relatively few useful restriction sites which could be used to clone the required DNA fragments. The Polymerase Chain Reaction (PCR) was, therefore, used to clone these genes. Using carefully designed primers (Appendix 1) incorporating unique restriction sites, the putative Dpol genes were selectively amplified and cloned. The first set of primers were designed to facilitate cloning of both K1 ORF 1 (primers CA1 and 2) and K2 ORF 2 (primers CA3 and CA4) into either pTTQ8 or 18. The primers were constructed such that an *Eco*R I site was located at the 5' end of the gene, immediately preceding of the second codon (AAA), and a *Sal* I site at the 3' end, just downstream of the termination codon. Cloning into the *Eco*R I-*Sal* I sites of the pTTQ polylinker, therefore, produced an in-frame translational with the *lacZ* gene, which in effect contributed only the initiating ATG, which was lost from the PCR product (Figure 3.1b).

The 2965 bp of K1 ORF 1 was amplified very selectively from pKLK111; there were no other major products. However, amplification of K2 ORF 2 from pKLK201 not only yielded the expected 2992 bp fragment, but also several additional products (Figure 3.1). The specificity of amplification was improved slightly by raising the annealing temperature (from 45^{o} to 60^{o}C, above which loss of the required product was observed) but some

Figure 3.1 The PCR amplification of K1 ORF 1 and K2 ORF 2, electrophoresed on a 0.8% w/v agarose gel. Lanes 1 & 12, *Hind* III digested lambda genome DNA size markers; lane 2, pKLK202, no enzyme control; lanes 4 & 5, pKLK202 amplifications; lane 6, pKLK111, no enzyme control; lanes 7 & 8, pKLK111 amplifications; lane 9, pKLK203, no enzyme control; lanes 10 & 11, pKLK203 amplifications. pKLK203 does not contain either K1 ORF 1 or K2 ORF 2.

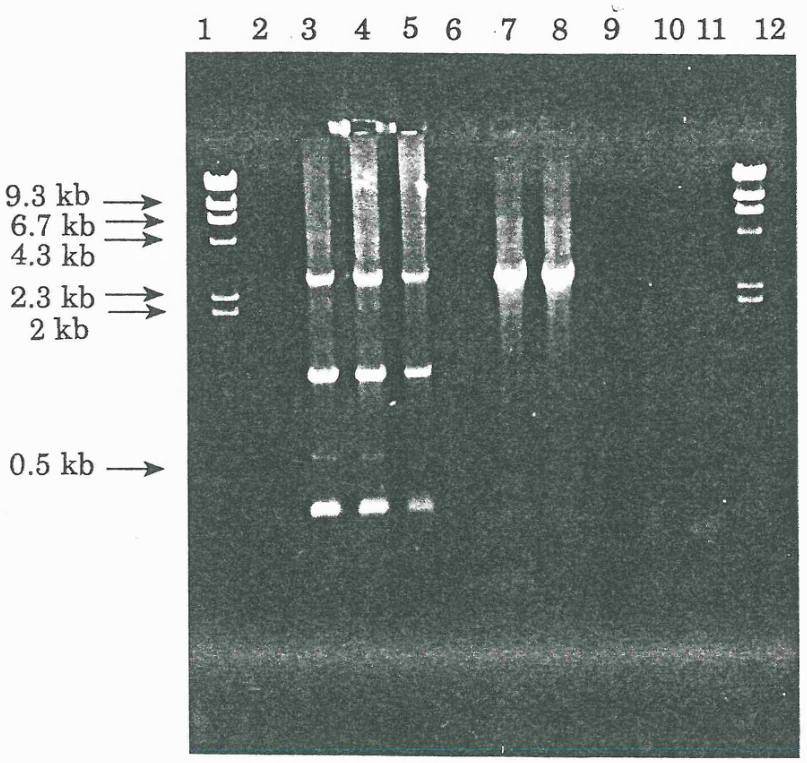


Figure 3.1b. The reading frame of the vector pTTQ8 and the junctions of the *lacZ* and *K1 ORF 1* and *K2 ORF 2* genes.

pTTQ 8

M N S R G S V D
GCG ATG AAT TCC CGG GGA TCC GTC GAC
*Eco*R I *Sma* I *Bam*H I *Sal* I

pTTQ8-*K1 ORF 1*

M *N* *S* *K* *D* *K*
GCG ATG *AAT TCC AAA GAT AAG*
*Eco*R I

pTTQ8-*K2 ORF 2*

M *N* *S* *E* *L* *A*
GCG ATG *AAT TCT GAA TTA GCA*
*Eco*R I

spurious amplifications remained. Computer analyses of the sequence of K2 for additional primer binding sites failed to reveal any other significant regions of homology, so the source of these additional products is not known. However, this was not a problem, as large amounts of the required product were synthesised and were simply gel isolated. Furthermore, this level of spurious product generation was not observed with later PCR amplifications using different primers.

3.2.1.1b Construction Of The pTTQ Recombinants.

PCR products were gel purified and digested with *Eco*R I and *Sal* I to yield the required cohesive ends. The pTTQ8 vector was digested with the same enzymes, gel purified and ligated with the restricted PCR products. The ligations were then used to transform *E. coli* and DNA from single transformants was screened by analysis of the electrophoretic mobility of the uncut, supercoiled plasmids. Restriction analysis of selected recombinants demonstrated the presence of the expected 3 kb PCR-generated fragments but also revealed that the parental plasmids appeared to migrate faster than expected in agarose gels. Further analysis confirmed the anomaly, the parental pTTQ8 vector was about 1 kb smaller than expected. This was probably due to a structural instability within the vector (Stark, Pers.Comm.; Maxwell, Pers.Comm.) as a result of a 54 bp duplication of the *lacI*^{Q} gene (bp 1668-1721 and 4119-4172). The Dpol fragments were, therefore, re-isolated from the pTTQ8 recombinants and cloned into a non-deleted pTTQ18 plasmid. These plasmids were named pCA1 (containing K1 ORF 1) and pCA2 (contained K2 ORF 2).

On a number of occasions incomplete restriction enzyme digestions at immediately adjacent sites led to difficulties obtaining the required recombinant plasmids. Therefore, to ensure that both sites within a polylinker were cleaved, digestions were always carried out independently and the DNA fragments gel purified. Similar problems were sometimes also encountered during the cloning of PCR products, by cleavage of terminal restriction sites incorporated *via* the primers. If such difficulties were experienced, PCR products were blunt end cloned into *Sma* I or *Eco*R V sites.

3.2.1.1c Expression From The pTTQ Vectors.

Analysis of IPTG induced cultures bearing pCA1 and pCA2, by SDS-PAGE of cell extracts, failed to reveal the presence of any high molecular weight (about 110 kDa) recombinant proteins. Using induction times of 1 to 4 h, IPTG concentrations varying from 0.1 to 0.5 mM, and examination of different cellular protein fractions, did not improve matters. No significant alterations in the culture growth rate were observed upon induction, in relation to uninduced cultures. Several independently isolated bacterial strains carrying plasmid clones pCA1 and pCA2 were analysed, and all gave the same results. The evidence therefore suggested that recombinant protein was not being synthesised. One of the reasons for this lack of expression may have been the inability of *E. coli* Rpol to efficiently transcribe this A+T rich DNA. The T7 Rpol vectors were therefore used to circumvent this problem.

3.2.1.2 The pET Vectors.

The T7 Rpol mediated expression system provided by the pET and pT7 vectors (Rosenberg *et al.*, 1987; Studier *et al.*, 1990) is particularly versatile in that it allows a more fundamental analysis of heterologous gene expression at both the *in vitro* and *in vivo* level. The presence of specific transcriptional initiation and termination signals allow transcripts to be synthesised *in vitro* using purified T7 Rpol. The strong T7 gene 10 promoter (ø10) provided in the pET vectors is that of the major capsid protein, which during T7 infection is synthesised at a rate much greater than any other viral or host protein. The translational start site (s10), the initiating ATG and termination signals (øT) are also those of gene 10. These T7 Rpol vectors have, for example, been used to express the ø29 Dpol (Bernad *et al.*, 1989 & 1990; Blasco *et al.*, 1990 & 1992; Soengas *et al.*, 1992) in *E. coli*.

T7 Rpol itself has a number of features which make it particularly useful for transcribing heterologous genes, both *in vivo* and *in vitro*, and which might allow expression of K-plasmid genes: it has a higher specific activity than *E. coli* Rpol (about 5 times more so); it is much more processive and can completely transcribe around pBR322 several times; efficient T7 Rpol terminator sequences appear to be rare in *E. coli*; it is also highly specific for its own promoter and does not initiate at any known *E. coli* promoters. In addition T7 Rpol is also resistant to the antibiotic rifampicin, which specifically acts on the β-subunit of Rpol

and prevents the formation of the first phosphodiester bond. Therefore, under conditions of rifampicin-induced inhibition of *E. coli* Rpol, any gene placed behind the T7 promoter should be selectively transcribed. Consequently this allows very sensitive detection of low level heterologous protein synthesis *in vivo*.

The *E. coli* host, BL21 (DE3) is deficient in the *lon* and *ompT* proteases which may facilitate stability of over-produced proteins. BL21(DE3) has a fragment of bacteriophage λ integrated into its chromosome which carries the *lacI*^{Q} gene and the T7 Rpol gene, regulated by the *lac*UV5 promoter. Transcription of the T7 Rpol gene is, therefore, exclusively IPTG inducible.

3.2.1.2a Modification Of Plasmid pET3a.

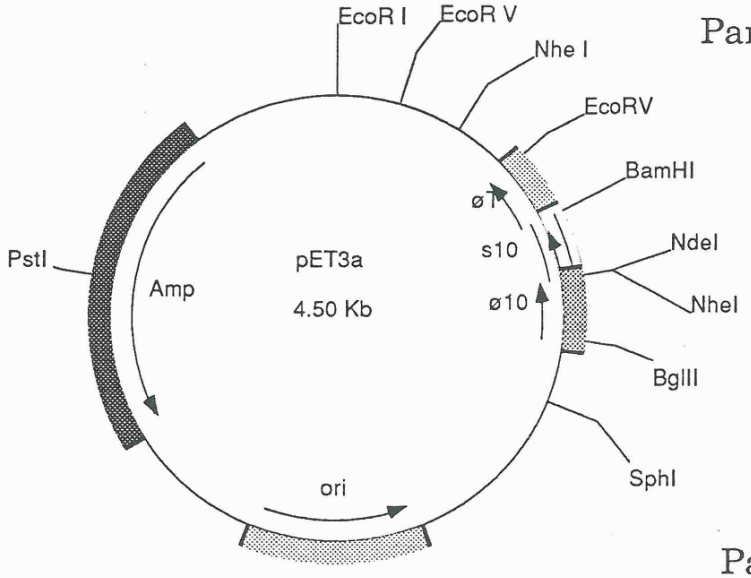
The putative Dpol gene of K2 ORF 2 could be cloned, in frame, into the pET3a vector (Figure 3.2A). However, in order to facilitate cloning, due to the presence of duplicate *Eco*R V and *Nhe* I sites, it was necessary to modify the plasmid prior to use. The plasmid was digested with *Eco*R V, gel isolated and religated. This deleted the small *Eco*R V fragment and effectively removed both the additional *Eco*R V and *Nhe* I sites (Figure 3.2). This plasmid was termed pCA8 and was equivalent to the pET11 series (Studier *et al*, 1990).

3.2.1.2b Construction Of pET Recombinants

Two new PCR primers, CA5 and CA6 (Appendix 1), with unique integral *Nhe* I and *Bam*H I restriction sites (at the 5' and 3'-termini, respectively) were designed. These were used to amplify the complete 3 kb sequence of K2 ORF 2 from pKLK201. The PCR product was digested with *Bam*H I and *Nhe* I and cloned into the corresponding sites in pCA8. Although the *Bam*H I-*Nhe* I digestion of pCA8 effectively removed most of the s10 sequence, the remainder of the upstream untranslated sequences (including the initiating ATG, the ribosome binding site and the critical spacing between this and the initiating codon) remained intact (Figure 3.2). Restriction analysis of selected clones confirmed the structure of the recombinant plasmid (pCA9) and this was then used to transform *E. coli* BL21 (DE3) for expression studies.

Figure 3.2 **Panel A.** Structure of the plasmid pET3a. Heterologous genes are transcribed by T7 Rpol, synthesised upon induction by the *E. coli* host BL21(DE3), with which this plasmid is used. Transcription of the heterologous gene is controlled by the T7 gene 10 promoter and terminator (ø10 & øT), and translational initiation is optimised by the presence of the 5' untranslated and extreme 5' translated region of gene 10. **Panel B.** The reading frame of the pET3a vector and the junction of the pCA8-K2 ORF 2 fusion in pCA9. **Panel C.** Modification of pET3a to yield pCA8, for use in expression of K2 ORF 2. The duplicated *Eco*R V and *Nhe* I sites were removed to facilitate cloning. A 10% w/v polyacrylamide gel showing *Eco*R V (lane 1, 3 & 5) & *Nhe* I (2, 4 & 6) digestions of pET3a (lanes 3 & 4) and two *Eco*R V deletion clones (lanes 1, 2, 5 & 6).

Panel A.



Panel B.

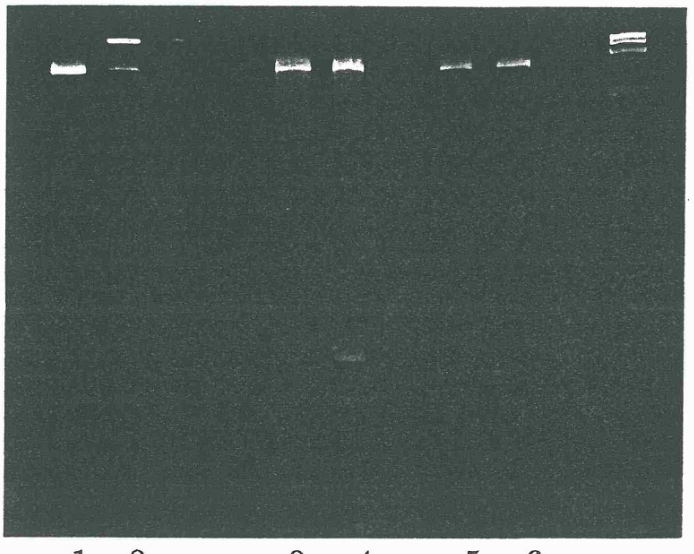
pET3a

M A S M T G G Q Q M G R G S
CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGC GGA TCC
Nde I *Nhe* I *Bam*H I

pCA8-K2 ORF2

M *A* *S* *M* *N* *D* *G*
CAT ATG *GCT AGC ATG AAT GAT GAA*
Nhe I

Panel C.



Over-expression of the complete Dpol genes, with their highly unfavourable A+T content and codon bias, was considered to pose a potential problem. Smaller fragments of these genes might, therefore be more amenable to the *E. coli* expression system. Both putative Dpols are complex molecules, but examination of their hydropathy profiles indicated that the extreme N-terminal section of K2 ORF 2 contained perhaps the single most continuously hydrophilic region of either protein. This portion might, therefore, be surface located and yield a relatively soluble product, and hence be the most suitable for over-expression. In addition, the extreme termini of proteins are often targeted for the synthesis of synthetic peptides as they tend to be surface located and relatively mobile (see Chapter 6). Consequently another primer, CA8 (Appendix 1), incorporating a unique *Bam*H I site was designed and used in conjunction with primer CA5, to amplify a 390 bp section at the extreme 5' end of K2 ORF 2 (Figure 3.9). The PCR product was gel purified, digested with *Bam*H I and *Nhe* I and ligated into *Bam*H I-*Nhe* I cleaved pCA8, to yield recombinant plasmid pCA13.

3.2.1.2c T7 RNA Polymerase Mediated Gene Expression.

Several control plasmids were used to check the ability of the system to express heterologous genes, as later recommended by Furlong *et al.* (1992). These were pET3a:*myo*I and pET3a:*sup*45 (section 2.2). Transformation of BL21 (DE3), and subsequent analysis, by SDS-PAGE, of crude soluble protein extracts from several induced cultures revealed that although not all plasmid-bearing strains were over-producers, some clones produced significant quantities of the expected 17 kDa protein. This is in agreement with the observations later reported by Furlong *et al.*, (1992) concerning the variability of the *E. coli* host. Upon induction of cultures containing several independently isolated clones of the recombinant Dpol plasmids pCA9 and pCA13 there was a pronounced reduction in culture growth rate, to a greater extent than observed with non-recombinant plasmid bearing strains. Extensive analysis of cellular protein fractions from induced cultures, however, failed to reveal any recombinant protein synthesis. Reductions in growth temperature, variations in induction time, and retransformation of freshly prepared competent BL21 (DE3) did not improve matters. Microscopic analysis of both pCA9 and pCA13-bearing induced cultures also did not reveal the presence of inclusion bodies.

3.2.1.2d Labelling Of T7 RNA Polymerase-Directed Expression Products.

As T7 Rpol is resistant to rifampicin (in contrast to *E. coli* Rpol) it is possible to induce the synthesis of T7 Rpol using host expression functions but to subsequently inhibit most host transcription. Any residual protein synthesis should, therefore, largely result from the translation of T7 Rpol transcribed genes. Low levels of recombinant protein might not be visible by SDS-PAGE analyses of crude cell extracts, even if cultures had been pre-treated with rifampicin. The pulse labelling of proteins synthesised after rifampicin treatment, using [^{35}S] methionine, might therefore, reveal the presence of the required product.

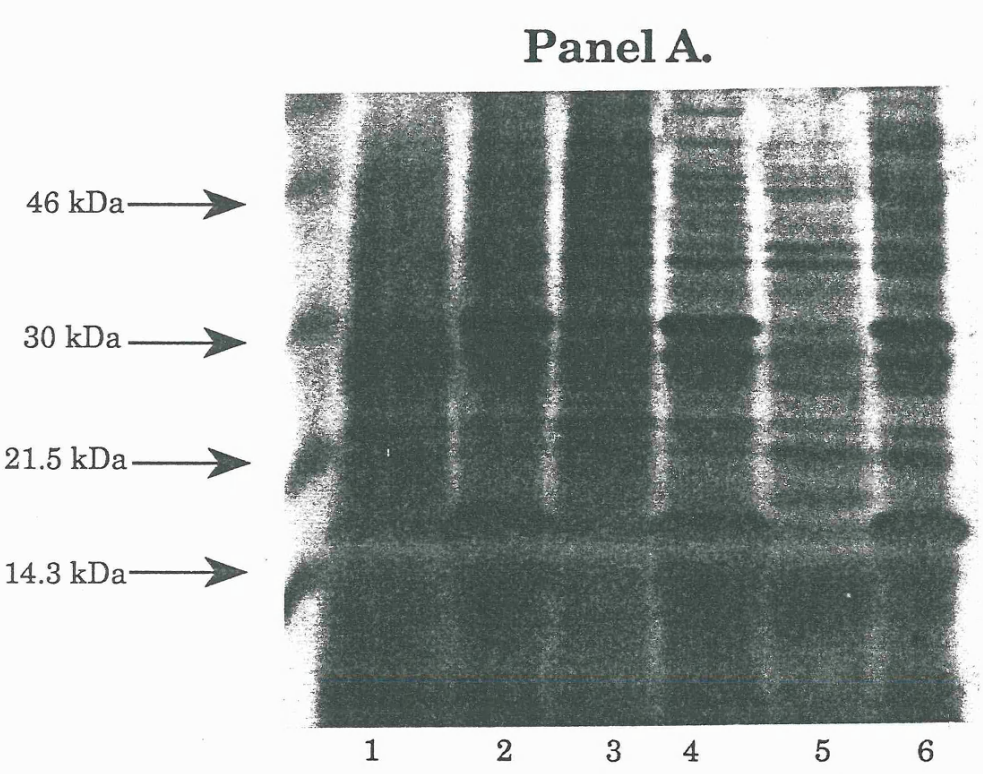
The incorporation of [^{35}S] methionine into the heterologous proteins encoded on plasmids pCA9 and pCA13 was analysed in parallel with the two control plasmids pET3a:*myo*I and pET3a:*sup*45. Significant quantities of the radiolabelled 14 and 49 kDa control peptides were observed upon induction of the control plasmids (Figure 3.3). Similar experiments using pCA9 and pCA13, however, failed to show any labelled recombinant protein synthesis (Figure 3.3). Here, although the absence of the expected 15 kDa product in Lanes 5 and 6 is not clear, numerous attempts were made to identify the K2 ORF2 product by both 15 and 20% SDS-PAGE. This suggested that either the mRNAs were not being translated, or that the translated or partially translated products were being very rapidly degraded.

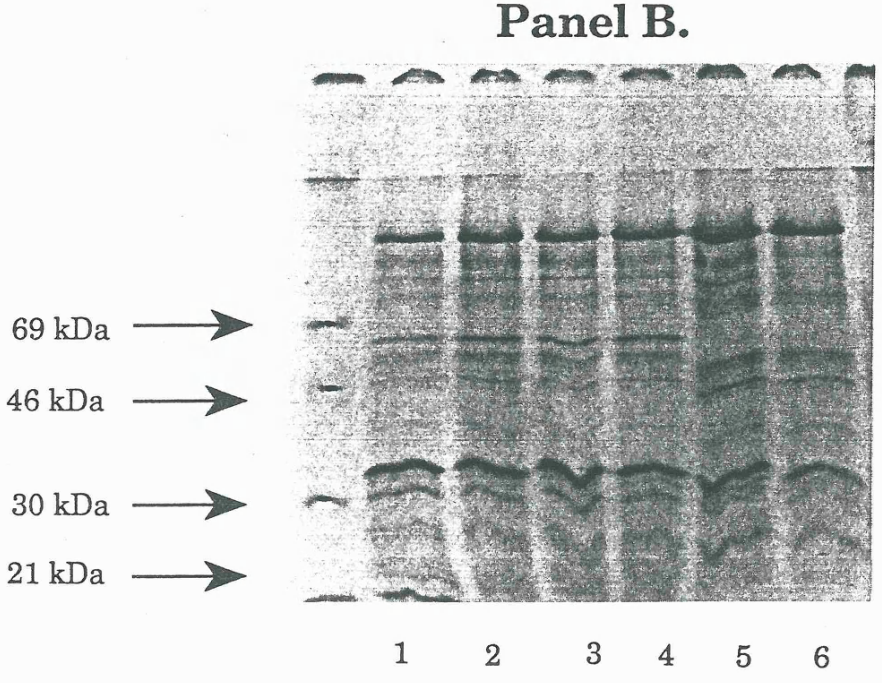
3.2.1.2e *In vitro* Transcription.

The inability of these plasmid systems to express the K-plasmid Dpol genes could be for a number of reasons. Some experimental evidence generated with yeast (Romanos & Boyd, 1988) had already suggested that heterologous systems might not be able to transcribe this DNA efficiently. Therefore, the size of the transcripts synthesised from the pET vector containing the complete K2 Dpol was analysed *in vitro*.

Analysis of transcripts synthesised *in vitro* from the control plasmid pET3a:*sup*45 demonstrated the production of the expected 1.5 kb mRNA. However, Northern analysis of RNA synthesised *in vitro* from pCA9, when probed with a radiolabelled fragment corresponding to the extreme 5' region region of the gene, revealed that no full length (3 kb) transcripts were being synthesised. The autoradiographs exhibited considerable smearing,

Figure 3.3 The radiolabelling of recombinant protein using the plasmid pET3a. **Panel A.** An autoradiograph of a 15 % w/v SDS-PAGE, showing the *in vivo* labelling of the 17 kDa pET3a:*myo*I product, using rifampicin inhibition of most host expression. Lanes 1, 3 & 5, uninduced cells; lanes 2, 4 & 6, IPTG induced cells. **Panel B.** An autoradiograph of a 10% SDS-PAGE obtained with the linked *in vitro* transcription and translation, using a cell-free *E. coli* extract. Lanes 1-4 are quadruplicate loadings showing the synthesis of the 50 kDa (approximately) product of the control pET3a: *sup*45 plasmid. Lanes 5 & 6 show samples produced using the same system with the plasmid pCA13, which contains the extreme N-terminal fragment of K2 ORF 2. Lanes 5 & 6 serve only to highlight the production of the *sup*45 product and not the expected 15 kDa K2 ORF2 product, visualisation of which would require analysis on a 15 or 20% gel.





but two major transcripts, of about 1 and 1.4 kb, were just discernible (Figure 3.4). Both these transcripts presumably initiate at the expected site, but terminated prematurely. Other transcripts, originating from internal initiations may have been present, but would not have been visualised using this probe. Increasing the proportion of rATP and rUTP in the transcription reactions also had no apparent effect on the size or amount of transcript produced. Furthermore, Northern analyses of mRNA transcribed *in vitro* from plasmid pCA13 demonstrated the synthesis of a single discrete messenger, of about 0.4 kb, but this appeared not to be translated.

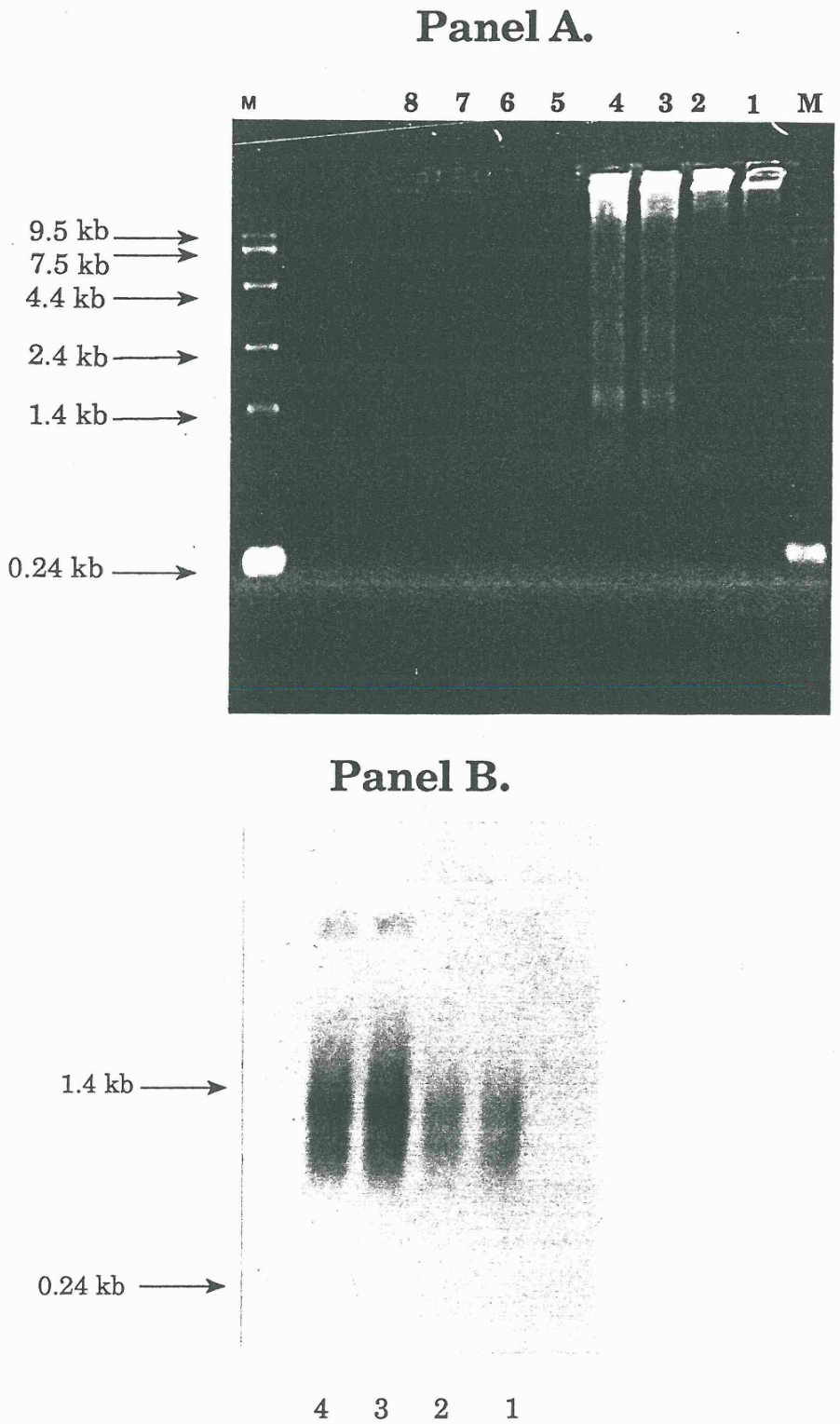
3.2.2 Expression Of The DNA Polymerase Genes As Fusion Proteins.

The expression of recombinant proteins as fusions proteins has a number of distinct advantages. The upstream gene sequence often provides strong transcriptional and translational initiation signals, allowing optimisation of these two rate limiting control points in *E. coli* gene expression. The conjugant protein moiety also often protects the recombinant peptide from proteolysis at either or both termini, depending upon the nature of the fusion, and may also facilitate isolation of the recombinant protein. Such an approach might, therefore, be successful with the K-plasmid Dpol genes.

3.2.2.1 The pRIT2T Vector.

This directs the expression of heterologous genes as C-terminal fusions with the *Staphylococcus aureus* cell wall component, protein A (Nilsson *et al.*, 1985). The protein A gene fusion system has been extensively developed, and has recently been reviewed by Nilsson & Abrahmsen (1990). In pRIT2T protein synthesis is directed by the λ P_r promoter. This is normally repressed by the λ cI gene product (responsible for maintaining λ lysogeny), but the isolation of a temperature sensitive allele (*ts* cI^{857}) facilitated the development of a heat induction system. Unfortunately the exact position and reading frames of the translational termination signals is not provided in the associated literature, so there are two possible sizes for the parental protein A fragment, depending upon the site of termination. Two non-sense codons lie immediately downstream of the polylinker, and a third further downstream. The parental peptide is actually a λ *cro*-protein A fusion containing the first 12 residues of the *cro* protein, plus 239 amino acids of protein A

Figure 3.4 Analysis of the RNA synthesised from plasmid pCA9 by T7 RpoI *in vitro*. **Panel A.** *In vitro* transcription reactions with the plasmid pCA9, which contains the 3 kb of K2 ORF 2 analysed on a denaturing 1.5% w/v agarose gel. Lanes 1-4, duplicate transcriptions; lanes 5 & 6, no plasmid control; lanes 7 & 8, no enzyme control; lanes M, RNA size markers. **Panel B.** Southern blot of the gel from Panel A, probed with a radiolabelled 5' fragment corresponding to the extreme 5' portion of K2 ORF 2.



upstream of the multiple cloning site. If the heterologous sequence contains a 3' stop codon or if translation terminates at the nonsense codons immediately adjacent to the polylinker, the protein A fragment is about 28 kDa. Termination at the site further downstream allows an additional 124 amino acids of protein A to be synthesised, resulting in a product of about 41 kDa. The protein A moiety consists of the IgG binding domains, and not only proteolytically stabilises the heterologous peptide, but also facilitates subsequent isolation from crude cell lysates, using IgG sepharose affinity chromatography.

3.2.2.1b Construction Of pRIT2T Recombinants.

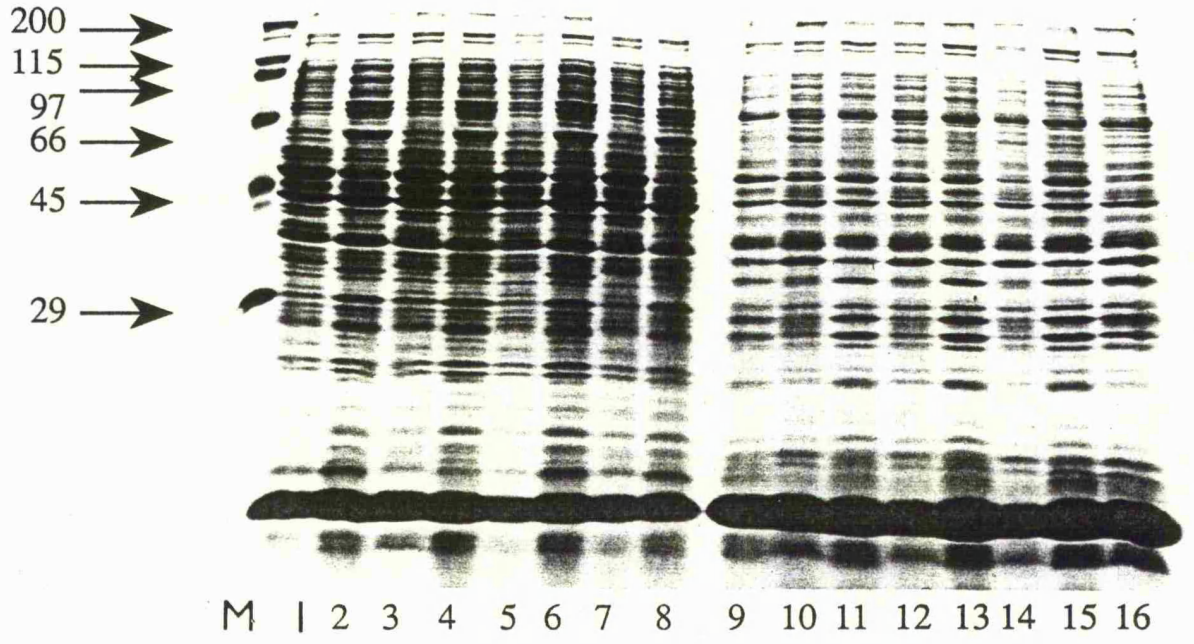
The reading frames and arrangement of restriction enzyme sites within the multiple cloning site of pRIT2T (that of pUC9) was identical to that in pTTQ18. The appropriate fragments were, therefore, simply sub-cloned from the pTTQ18 recombinants. The K1 ORF 1 and K2 ORF 2 fragments generated by PCR, were released from pCA1 and pCA2 by *Eco*R I-*Sal* I digestion and sub-cloned into the *Eco*R I-*Sal* I sites of pRIT2T, yielding plasmids pCA3 and 4. Plasmids were constructed in *E. coli* NM522, but were subsequently transferred to *E. coli* N4830-I (which carries the *ts* cI^{857} gene) for expression studies.

3.2.2.1c Over-Expression From pRIT2T Recombinants.

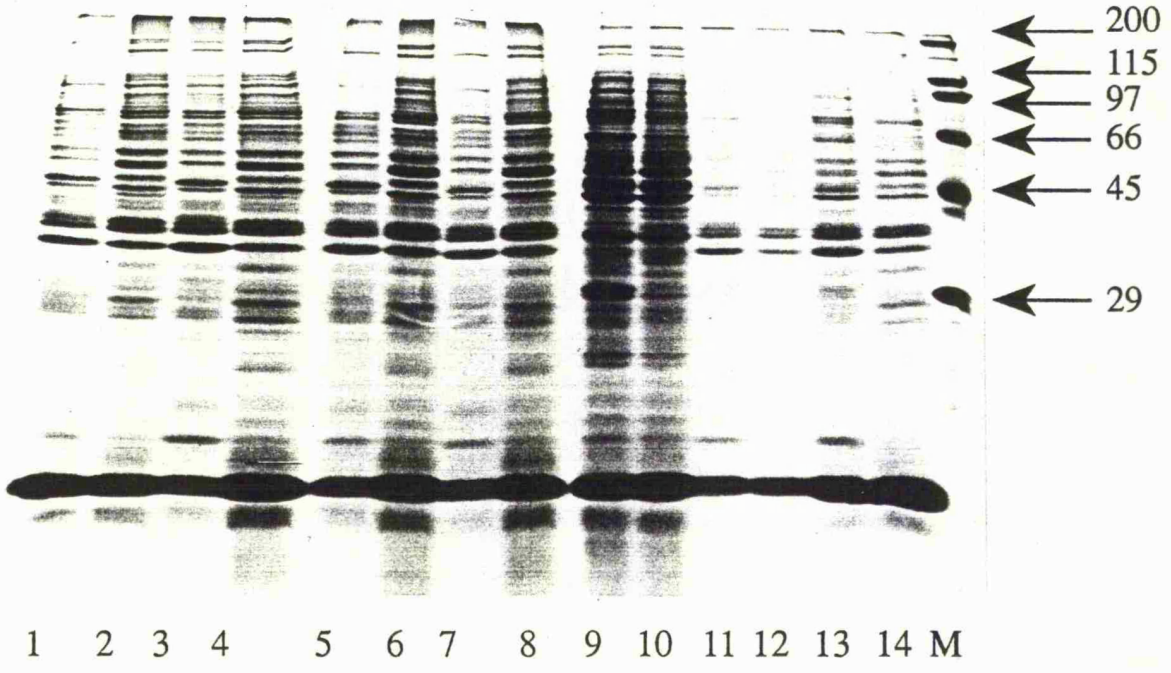
As both the putative Dpol genes were cloned into pRIT2T with their endogenous 3' nonsense codons, these together with the protein A fragment should have resulted in fusions proteins of approximately 140 kDa. However, extensive analysis, by SDS-PAGE, of extracts from several independently isolated clones containing either pCA3 or pCA4 failed to show any trace of the expected high molecular weight products (Figure 3.5). Phase contrast light microscopic examination of induced cells also failed to reveal the formation of inclusion bodies. Surprisingly, analysis of crude cell extracts from the induced parental pRIT2T control did not, generally, show significant amounts of the expected non-fusion protein A, although this may be the identity of the 30 kDa protein in lane 9 of Figure 3.5. Nevertheless, IgG sepharose affinity chromatography of soluble protein extracts prepared from induced strains containing pRIT2T yielded disappointingly small amounts of the parental peptide. Affinity chromatography of the induced recombinant plasmid bearing cells gave similar results, with low levels of non-specifically bound or degraded products. The

Figure 3.5. 15% SDS-PAGE of various solublised cell fractions prepared from *E. coli* N4830-1 containing the pRIT2T recombinants pCA3 and pCA4. **Panel A.** Lanes 1-8 are sample A, and lanes 9-16 are sample B (prepared according to the method outlined in Section 2.6.1.1a) of pCA3 (1-4 & 9-12) and pCA4 (5-8 & 13-16). Lane M is molecular weight markers (in kDa). Lanes 1, 3, 9 & 11 are from uninduced pCA3; Lanes 2, 4, 10 & 12 are induced pCA3; lanes 5, 7, 13 & 15 are from uninduced pCA4; lanes 6, 8, 14 & 16 are induced pCA4. **Panel B.** Lanes 1-8 are sample C of pCA3 (1-4) and pCA4 (5-8). Lanes 1 & 3, pCA3, uninduced; lanes 2 & 4, pCA3, induced; lanes 5 & 7, pCA4, uninduced; lanes 6 & 8, pCA4, induced. Lanes 9-14 are samples prepared from *E. coli* containing the parental pRIT2T vector. Lanes 9 & 10, sample A, induced & uninduced; lanes 11 & 12, sample D, induced & uninduced; lanes 13 & 14, sample C, induced and uninduced.

Panel A.



Panel B.



heat induction system employed by these vectors, is now thought to be much less specific than previously suspected. Such a temperature shift also appears to induce heat shock proteins, some of which are proteases (eg. the *lon* protease). The lack of parental protein A isolated by affinity chromatography may, however, have been due to denaturation of the IgG sepharose, which was accidentally stored at -20°C, rather than the recommended 4°C.

3.2.2.2 The pRIT5 Vector.

The results generated thus far implied that *E. coli* might not be a particularly suitable host for high level expression of K-plasmid genes. An examination of the nt and codon composition of the K-plasmids, *E. coli*, *S. cerevisiae* and *S. aureus* (Table 3.1) suggested that one possible alternative was expression in *S. aureus*.

pRIT5 is an *E. coli* - *S. aureus* shuttle-expression vector. Like pRIT2T, heterologous genes are expressed as C-terminal fusions with protein A and can be purified in the same manner. However, gene expression in this case is directed by the constitutive protein A promoter, and hence expression cannot be regulated. In *E. coli* fusion proteins are directed to the periplasm, but in *S. aureus* are secreted.

3.2.2.2a Construction Of pRIT5 Recombinants.

The polylinker in the plasmid pRIT5 was also identical to that in pTTQ18. This facilitated the construction of pRIT5 recombinants by simple sub-cloning of the 3 Kb *Eco*R I-*Sal* I fragments of K1 ORF1 and K2 ORF2 from pCA1 and 2. However, numerous attempts to clone these DNA fragments failed. Transformation frequencies were consistently low (3-$6x10^3$/μg, compared to $1x10^6$ to $1x10^7$ transformants/μg control plasmid) and analysis of the resultant ampicillin resistant colonies (by *Pvu* II cleavage within the ampicillin resistance gene) revealed small deleted plasmids of approximately 2 kb. This implied that the presence of the K-plasmid Dpol genes in pRIT5 is deleterious to the host cells and resulted in the 'selection' of cells carrying deleted variants of these plasmids. As pRIT5 directs heterologous genes to the periplasm in *E. coli*, the toxic effect of the gene fusion may reflect the constitutive synthesis and insertion of low levels of the product into the cell membrane.

Table 3.1. Comparison of the Codon Bias, In Terms of the RSCU in the K-plasmids, *S. cerevisiae, E. coli* and *S. aureus.* See Section 3.3 for explanation.

Amino acid	Codon	RSCU				Amino acid	Codon	RSCU			
		E..coli	K-plasmid	*S.cerevisiae*	*S.aureus*			*E.coli*	K-plasmid	*S.cerevisiae*	*S.aureus*
Phe (F)	UUU	0.46	1.54	0.3	1.25	Thr (T)	ACA	0.14	1.72	0.24	1.88
	UUC	1.54	0.5	1.7	0.75		ACG	0.19	0.1	0.12	0.29
Leu (L)	UUA	0.11	3.95	0.68	3.47	Ala (A)	GCU	1.88	2.1	2.78	1.72
	UUG	0.11	0.4	4.77	0.87		GCC	0.23	0.17	1.13	0.16
	CUU	0.23	0.7	0.13	0.6		GCA	1.1	1.61	0.06	1.81
	CUC	0.2	0.07	0.03	0.14		GCG	0.8	0.12	0.04	0.31
	CUA	0.04	0.63	0.39	0.69	Tyr (Y)	UAU	0.39	1.69	0.27	1.68
	CUG	5.33	0.250	0.01	0.14	His (H)	UAC	1.61	0.31	1.73	0.33
Isoleu (I)	AUU	0.47	0.92	1.45	1.35		CAU	0.45	1.77	0.36	1.57
	AUC	2.53	0.08	1.53	0.59		CAC	1.55	0.23	1.64	0.43
	AUA	0.008	2.0	0.03	1.1	Gln (Q)	CAA	0.22	1.67	1.84	1.89
Met (M)	AUG	1	1	1	1		CAG	1.78	0.33	0.16	0.11
Val (V)	GUU	2.24	1.63	2.25	1.87	Gly (G)	GGU	2.29	1.71	3.75	2.08
	GUC	0.15	0.17	1.68	0.44		GGC	1.65	0.11	0.14	0.8
	GUA	1.11	1.98	0.02	1.3		GGA	0.02	2.03	0.07	0.85
	GUG	0.5	0.22	0.05	0.38		GGG	0.04	0.15	0.05	0.28
Ser (S)	UCU	2.6	2.62	3	1.68	Asn (N)	AAU	0.1	1.72	0.1	1.23
	UCC	1.91	0.27	2.61	0.27		AAC	1.93	0.27	1.9	0.77
	AGU	0.22	1.68	0.13	1.44	Lys (K)	AAA	1.6	1.7	0.38	1.66
	AGC	1.1	0.16	0.04	1.05		AAG	0.4	0.3	1.62	0.34
	UCA	0.2	1.14	0.18	1.32	Asp (D)	GAU	0.61	1.73	0.66	1.46
	UCG	0.04	0.13	0.04	0.23		GAC	1.4	0.27	1.34	0.54
Pro (P)	CCU	0.23	2.32	0.64	1.98	Glu (E)	GAA	1.59	1.75	1.89	1.67
	CCC	0.04	0.21	0.05	0.18		GAG	0.41	0.25	0.11	0.33
	CCA	0.44	1.35	3.31	1.59	Cys (C)	UGU	0.67	1.75	1.9	2
	CCG	3.3	0.11	O	0.25		UGC	1.33	0.25	0.1	0
Thr (T)	ACU	1.8	2	1.84	1.69	Arg (R)	CGU	4.38	0.18	0.3	1.65
	ACC	1.87	0.23	1.8	0.13		CGC	1.56	0.05	0	0.71
	ACA	0.14	1.72	0.24	1.88		CGA	0.02	0.13	0	0.35
	ACG	0.19	0.1	0.12	0.29		CGG	0.02	0	0	0.12
	ACU	1.8	2	1.84	1.69		AGA	0.02	5.32	5.54	3.06
	ACC	1.87	0.23	1.8	0.13		AGG	0.01	0.33	0.17	0.12
						Trp (W)	UGG	1	1	1	1

The lack of a regulated expression system for *S. aureus* effectively halted this work. However, should such plasmids become available it would be interesting to experiment further with this system.

3.2.2.3 The pGEX Vectors.

The failure of the pRIT plasmids to yield significant quantities of fusion protein was disappointing. However, the non-fusion protein A did not appear to be synthesised particularly efficiently from the parental plasmids. In a final attempt to express the native K-plasmid genes in *E. coli*, an alternative gene fusion system was tried.

The pGEX plasmids (Smith & Corcoran, 1987; Smith & Johnson, 1988) direct the production of heterologous proteins in *E. coli* as C-terminal cytosolic fusions with the 27 kDa *Schistome japonicum* glutathione-S-transferase (GST). Heterologous gene expression is regulated by the *tac* promoter and the presence of the *lac*I^Q gene ensures tight repression in the absence of IPTG induction. Suitable restriction sites are available in all three reading frames (Figure 3.6), and soluble fusion products can be purified, in a single step, by glutathione affinity chromatography. Furthermore, two of the three vectors (pGEX2T and pGEX3X) yield products which contain engineered proteolytic cleavage sites. The N-terminal GST moiety can therefore be removed after purification, facilitating the application of proteins synthesised using these vectors, in antibody generation.

The majority (75%) of proteins (ranging in size from 43 to 67 kDa) expressed in these plasmids by the authors, were reported to be soluble and could be isolated by affinity chromatography. The failure in the minority of cases they attributed to production of insoluble fusions. Problems were experienced with proteins of over 50 kDa, containing strong hydrophobic or highly charged sequences. This has since also been observed by Kaelin *et al.*, (1991), but these plasmids have recently been used with some success by Defeo-Jones *et al.*, (1991) and Driscoll *et al.*, (1991).

3.2.2.3a Construction Of pGEX Recombinants.

The 1.7 kb *Pst* I-*Sal* I fragment of K1 ORF1 (lacking the 5' 1.3 kb of the gene, Figure 3.9) from plasmid pCA1 was sub-cloned into *Pst* I-*Sal* I digested pIC19R (Marsh *et al.*, 1984).

Figure 3.6 The structure of the two pGEX plasmids used in this study, showing the cloning sites and the reading frames.

pGEX1

```
 P   K   S   D   P   R   E   F   I   V   T   D  Stop
CCA AAA TCG GAT CCC CGG GAA TTC ATC GTC ACT GAC TGA
          BamH I  Sma I  EcoR I
```

pGEX2T

```
 P   K   S   D   L   V   P   R   G   S   P   G   I   H   R   D  Stop
CCA AAA TCG GAT CTG GTT CCG CGT GGA TCC CCG GGA ATT CAT CGT GAC TGA
                      BamH I Sma I EcoR I
```

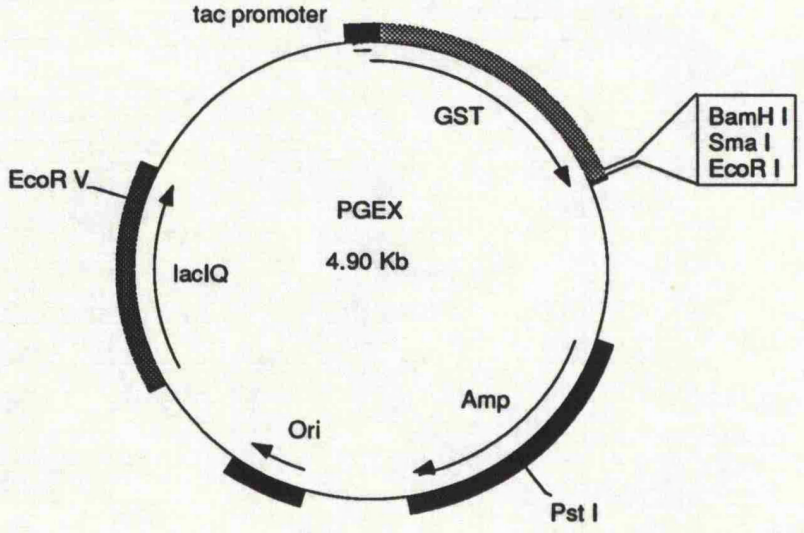


Figure 3.6b. The structure of the junction between the GST peptide in pGEX1 and the inverted 470 bp *Bgl II-Sau* 3a fragment of K2 ORF2. This illustrates the presence of a stop codon very close to the beginning of the ORF2 sequence.

```
                D   L   L   N   I   L  Stop
pGEX1   G GAT CTT TTA AAT ATA TTA TAG  K2 ORF2
        C CTA GAA AAT TTA TAT AAT ATC
          BamH I/Sau 3a
```

Recombinant plasmids were identified by insertional inactivation of the *lacZ* product on selective media supplemented with X-gal and IPTG. Clones containing the required DNA fragment were unable to cleave the chromogenic substrate X-gal and the colonies therefore remained white. Restriction analysis confirmed the structure of a recombinant plasmid, which was designated pCA5. An *Eco*R I-*Bgl* II digest of pCA5 then released a fragment which was cloned into the *Eco*R I-*Bam*H I site of pGEX2T, producing an in-frame translational fusion with the GST moiety (*Bgl* II and *Bam*H I digestions produce compatible cohesive ends). This plasmid was named pCA6.

An *Eco*R I site exists naturally just downstream of the 3' end of K2 ORF 2. A *Bgl* II-*Eco*R I digest of pKLK201 released a 2.1 kb fragment (lacking the 5' 1 kb of the native gene, Figure 3.9), which was cloned directly into the *Bam*H I-*Eco*R I sites of pGEX1, producing plasmid pCA7.

By this stage, the probability of expressing large fragments of the K-plasmid genes was considered to be remote. Therefore, attempts were also made to express smaller fragments of K2 ORF 2 in the pGEX vectors (Figure 3.9). The 2.1 Kb *Bgl* II - *Eco*R I fragment from pKLK201 was gel purified and then digested with *Sau* 3a. This yielded 3 fragments of 470 (the most 5' fragment), 640 and 940 bp (the most 3' fragment). The 470 bp *Bgl* II-*Sau* 3a fragment was cloned into *Bam*H I cut pGEX1, to yield plasmid pCA10. As both *Bgl* II and *Sau* 3a cohesive termini are compatible with *Bam*H I sticky ends there was no control over the orientation of insertion. Transformants were screened for the correct orientation using a convenient internal *Eco*R V site which lies close to the *Bgl* II site in the 470 bp fragment. Clones containing the fragment in the reverse orientation were termed pCA14. The 940 bp *Sau* 3a-*Eco*R I fragment, corresponding to the extreme 3' end of the gene (Figure 3.9) was cloned into *Bam*H I-*Eco*R I cut pGEX1, to give plasmid pCA11.

As previously discussed (Section 3.2.1.2b), examination of the hydrophobicity profile of K2 ORF 2 revealed that the single continuously most hydrophilic portion of the protein was at the extreme N-terminus. The PCR, using primers CA9 and CA10 was therefore used to clone the terminal 390 bp region of ORF 2, yielding a fusion of the first 130 amino acids with the GST carrier. Unique *Bam*H I and *Eco*R I restriction sites were incorporated within

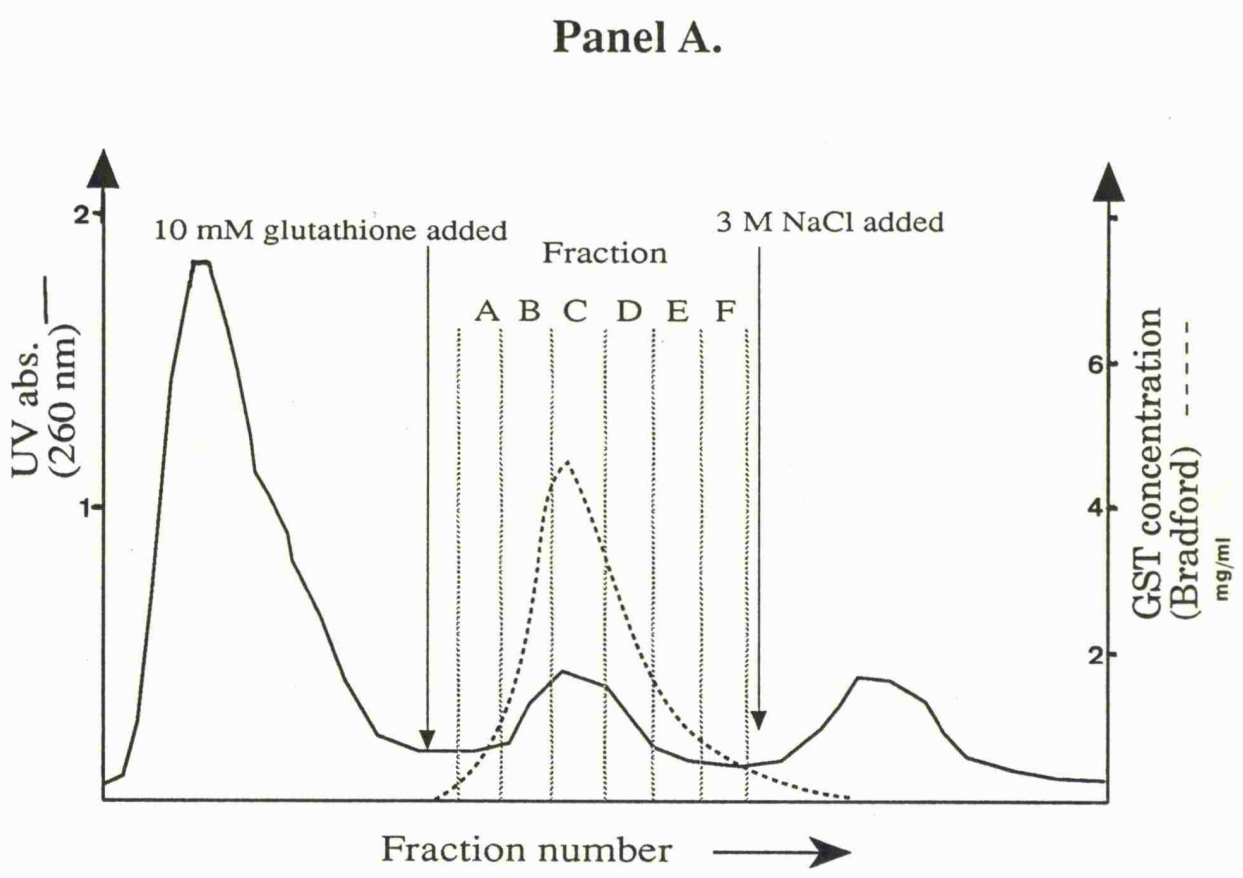
the PCR primers to allow direct cloning of the product, in frame into pGEX2T (to give plasmid pCA12).

3.2.2.3b Expression From GST Vectors.

Solublised cell extracts from *E. coli* containing the parental pGEX1 or pGEX2T vectors were analysed for their GST over-production, prior to any attempts to detect fusion proteins. A typical protein elution profile is shown in Figure 3.7. A relatively large peak was observed to elute after addition of the glutathione and the resultant fractions were analysed by quantitative protein assay and SDS-PAGE. A calibration curve using an ovalbumin stock was constructed from which the GST concentrations were calculated. Column chromatographic fractions from the pGEX1 and pGEX2T controls contained considerable quantities of a purified 27.5 kDa protein, as revealed by SDS-PAGE analyses and protein quantitation (Figure 3.7 and 3.8). A specific GST enzyme assay was then used to confirm the identity of the protein in the eluates. GST catalyses the reaction of glutathione with 1-chloro-2,4-dinitrobenzene, and results in the formation of a coloured product which absorbs at 340 nm. This demonstrated that the eluate fractions with the high protein concentrations also contained the greatest GST activity.

The expected sizes of the fusion products from pCA6 and pCA7 were approximately 82 and 92 kDa, respectively. Affinity chromatography, protein quantitation and SDS-PAGE of proteins extracted from several independently isolated clones revealed that, in both cases, very little recombinant protein was produced. Only very small peaks in UV absorption were observed following the addition of glutathione to the affinity column. Protein quantification of eluates from 500 ml cultures (Table 3.2) confirmed these very low levels, demonstrating that the majority of the fractions were almost completely devoid of protein. Enzyme assays also failed to detect any GST activity. Several very faint bands were visible by SDS-PAGE analysis of fractions with the 'highest' protein concentrations, but these appeared to be primarily due to non-specific binding or degradation products. Analyses of strains bearing the recombinant plasmids containing small K2 ORF 2 fragments (pCA12) yielded similar results. No expression of recombinant protein was observed during extensive analyses of various cellular proteins (including insoluble fractions) or affinity chromatographic elution

Figure 3.7 **Panel A.** A typical protein elution profile for the analysis of soluble GST fusion products, produced using the pGEX vectors. The profile below was obtained with the control plasmid pGEX1. **Panel B.** The protein content, as revealed by 10% SDS-PAGE, of fractions C (lanes 2 & 3) and D (lanes 4 & 5). Lanes 1 & 6 are protein molecular weight markers (kDa).



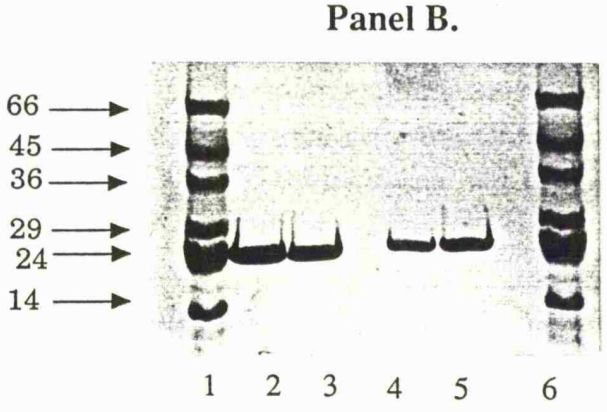
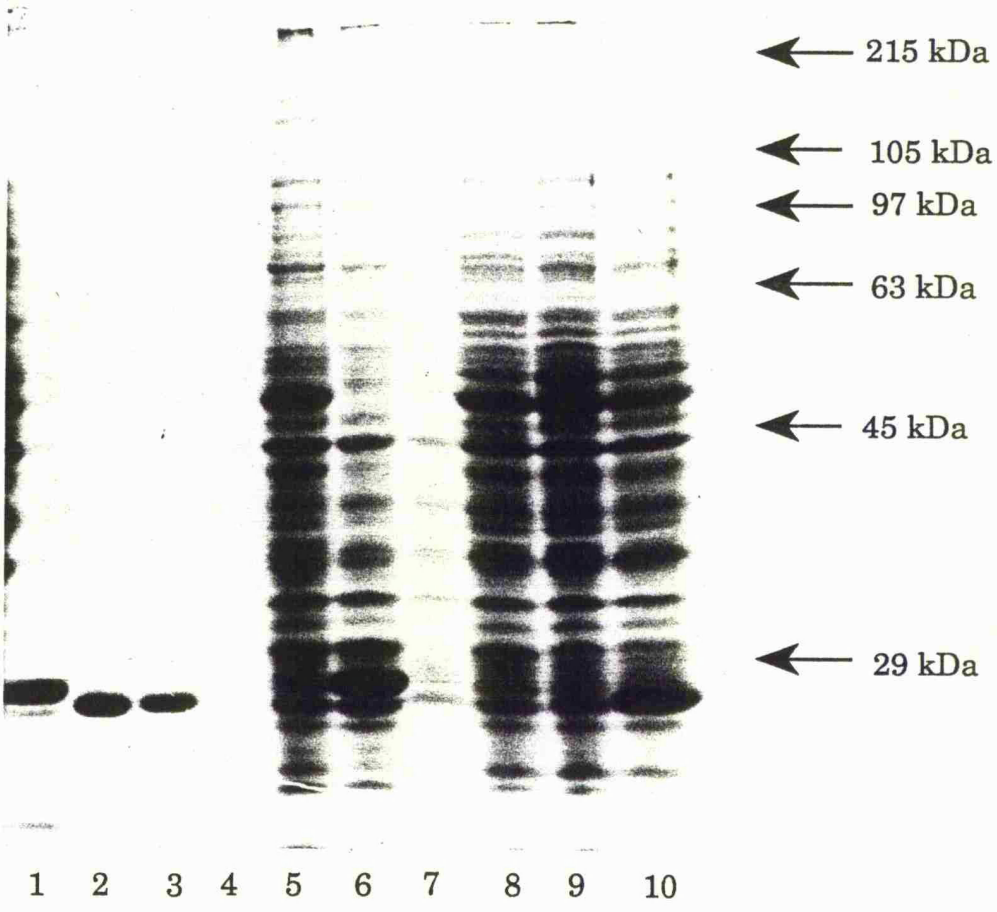


Table 3.2 Quantitation of the protein eluted by the addition of reduced glutathione from the glutathione affinity column, following the application of soluble protein samples prepared from the recombinant plasmids pCA6 and pCA7. Protein concentrations were calculated using the Biorad protein assay, with a BSA calibration curve.

Fraction	OD_{595nm}	[protein] (μg/ml)
pCA6		
2	0.082	40
3	0.189	100
4	0.122	60
5	0.057	25
6	0	0
pCA7		
6	0.011	7
7	0.186	100
8	0.526	400
9	0.595	480
10	0.412	400
11	0.212	120
12	0.129	80
13	0.084	40
14	0.051	20

Figure 3.8 15% SDS-PAGE of soluble protein fractions prepared from pGEX plasmids, pCA10, pCA14 and pGEX1 in *E. coli* NM522. Both pCA10 and pCA14 contain the identical 470 bp *Bgl* II-*Sau* 3a internal fragment of K2 ORF 2, cloned into the *Bam*H I site of pGEX1. However, the two plasmids differ in the orientation of this fragment; pCA14 has the K2 ORF 2 fragment inserted in the wrong orientation. This was predicted to result in the accumulation of parental GST, due to the occurence of in-frame non sense codons, in contrast to the situation with pCA10. Lane 1-4, affinity purified GST produced from plasmid pCA14 (1), pGEX1 (2 & 3), and pCA10 (4). Lanes 5-10 show total cellular protein prepared from *E. coli* strains containing uninduced (5) & induced (6) pCA14, induced (7) & uninduced (8) pCA10, uninduced (9) and induced (10) pGEX1.



fractions. Varying the induction time from 1 to 18 h, or the induction rate (the IPTG concentration and the culture growth temperature) did not significantly alter the protein production profile of these recombinants.

The complete absence of the parental GST peptide was immediately obvious in all cases. That this was due to the nature of the heterologous DNA (or its product) cloned into the vectors was demonstrated by the insertion of the 470 bp *Bgl* II-*Sau* 3a of K2 ORF 2 in the reverse orientation in pGEX1 (plasmid pCA14). This effectively produced an out of frame translational fusion. Examination of the appropriate DNA sequences revealed that a non-sense codon was encountered very close to the end of the GST sequence (Figure 3.6b), and would be expected to result in the production of the parental GST moiety. Upon induction of these strains, large amounts of the soluble 27 kDa parental GST were observed, both by analysis of crude cell extracts and affinity chromatography (Figure 3.8). Evidence therefore suggests that the coupling of the K-plasmid DNA to the carrier peptide resulted in destabilisation of the fusion products, at transcriptional and/or translational levels.

3.3 DISCUSSION.

Evidence to date suggests that the linear "killer" plasmids of *K. lactis* probably replicate using a mechanism similar to that employed by other linear ds DNA plasmids. Such mechanisms have been particularly well characterised in Ad, ø29 and PRD1. In these cases elucidation of the minimal requirements for DNA replication was facilitated by the cloning and over-expression of the genes encoding the essential replicatory proteins.

The ultimate objective of work in our laboratory is to establish a minimal system which could support the replication of the K-plasmids. Isolation of the native K-plasmid replicative enzymes from cytosolic extracts of *K. lactis* using conventional biochemical purification techniques, would be difficult due to the natural low abundance of these proteins. Furthermore, specific and sensitive assays need to be available to aid purification. Initial steps to develop such assays, based upon the ability to distinguish K-plasmid encoded Rpol or Dpol activities from the host encoded activities proved unsuccessful (Brophy & Meacock, unpublished results). Similar problems hampered the isolation of Ad replicatory proteins for many years. A much more exciting prospect, was to employ recombinant DNA technology.

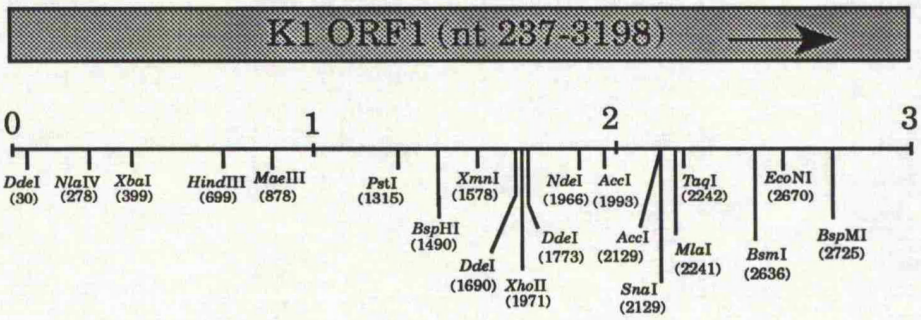
The rationale behind our strategy was, therefore, similar to that successfully employed by other researchers, and was based upon the over-expression of K-plasmid ORFs in the most versatile expression system currently available, that of *E. coli*. The ø29 proteins p2, p3, p5 and p6, the Ad pTP and Dpol, and the PRD1 TP and Dpol have all been over-expressed in *E. coli*. Production of an enzymically functional recombinant protein would, of course, have been the ideal case. However, over-expression of gene fragments and purification of the recombinant protein would be sufficient to allow the generation of polyclonal antibodies, which could subsequently be used to detect, isolate and characterise the native protein from crude yeast extracts. There were then numerous aspects of Dpol functioning, particularly the 'linear plasmid-specific' properties, which could have been investigated.

Despite the widespread success of recombinant DNA techniques, the failure of these systems to efficiently over-produce certain proteins has, however, been poorly documented. The results outlined in the preceding sections have demonstrated that difficulties have recently been experienced in the over-expression of the putative K-plasmid Dpol genes in *E. coli*. These results are summarised in Figure 3.9. Experimental evidence suggests that these problems have arisen due to both transcriptional and translational factors.

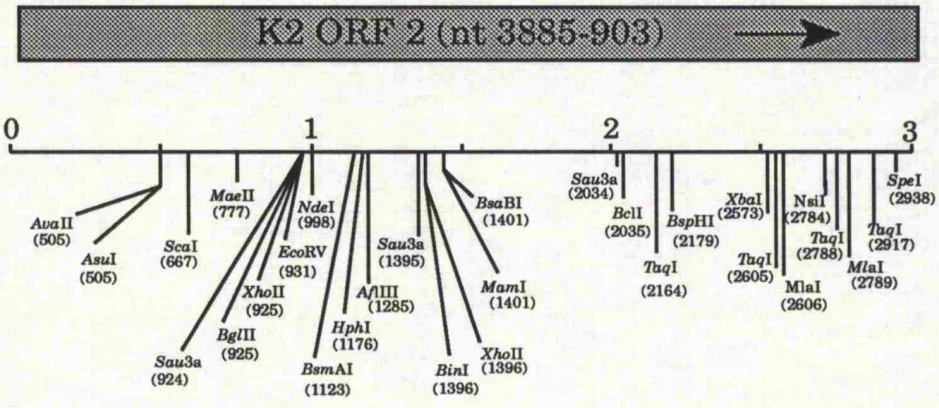
Like other extranuclear genetic elements, including other linear plasmids (see Table 1.4), the K-plasmids have an extraordinarily high A+T content (that of the K1 and K2 Dpol genes is 73.9% and 75.4%, respectively). This is probably the fundamental cause of the problems which have been experienced, and the effects are probably manifest in both transcription and translation.

At the transcriptional level, the occurrence of fortuitous transcriptional initiation and termination sequences, which are recognised by cellular Rpols is a distinct possibility. Such observations have previously been reported in this (Romanos & Boyd, 1988) and other systems. Northern analyses of transcripts produced from K1 ORFs in yeast nuclear vectors demonstrated that in all cases, except ORF 3, no native transcripts were detected. Abundant aberrant transcripts corresponding to ORFs 1, 2 and 4 were detected. Transcripts derived from ORF 2 fell into two broad bands of about 0.5 and 0.9 kb; each containing several

Figure 3.9 A diagramatic summary of the regions of the K-plasmid Dpol genes which have been cloned in *E.coli* expression vectors, and the results of subsequent expression studies.



Plasmid			Expression?
pCA1	*Eco*RI	3 kb PCR fragment in pTTQ8 — *Sal*I	NO
pCA3	*Eco*RI	3 kb PCR fragment in pRIT2T — *Sal*I	NO
pCA6	*Pst*I	1.7 kb PCR fragment in pGEX2T — *Sal*I	NO



Plasmid			Expression?
pCA2	*Eco*RI	3 kb PCR fragment in pTTQ8 — *Sal*I	NO
pCA4	*Eco*RI	3 kb PCR fragment in pRIT2T — *Sal*I	NO
pCA9	*Nhe*I	3 kb PCR fragment in pCA8 — *Bam*HI	NO
pCA12	*Eco*RI — *Bam*HI	390 bp PCR product in pGEX2T	NO
pCA13	*Nhe*I — *Bam*HI	390 bp PCR product in pCA8	NO
pCA7	*Bgl*II	1.9 kb PCR fragment in pGEX1 — *Eco*RI	NO
pCA10 & 14	*Bgl*II — *Sau*3a	470 bp native fragment in pGEX1	NO
	*Sau*3a	900 bp native fragment in pGEX1 — *Eco*RI	NO
pCA30	*Eco*RI — *Bam*HI	160 bp resynthesized fragment in pGEX2T	YES
pCA31	*Bgl*II — *Bam*HI	160 bp resynthesized fragment in pET3a	NO

distinct species corresponding to internal sections of the gene. All transcripts from ORF 1 were of less than 0.5 kb. These results support the view that K-plasmid ORFs are not efficiently transcribed by yeast nuclear Rpols and that transcription probably arose due to the presence of sequences which fortuitously acted as transcriptional initiators and terminators. This also suggests that expression of ORF 3 reported by Tokunaga *et al.* (1987) was purely fortuitous. Presumably there was a sequence upstream of ORF 3 in this construction that allowed a yeast Rpol to initiate and subsequently transcribe this relatively small (750 bp) gene. Romanos (1992) listed six examples where the high A+T content of a particular gene caused reduced gene expression in *S. cerevisiae*. The Tetanus toxin C fragment of *Clostridium tetani* (71% A+T) produced truncated mRNAs due to the presence of at least six fortuitous transcriptional terminators. These were eliminated by increasing the G+C content of the gene (from 29 to 47%); but full length transcript was only detected in the fully synthetic gene.

The high A+T content is also likely to mean that the mRNAs are highly structured. 'Stem-loop', 'Fold' and 'Terminator' analyses (using the University of Wisconsin Molecular Biology Software Package) of both K1 ORF 1 and K2 ORF 2 mRNAs reveal the presence of many regions with the potential to form higher order structuring. The Stem-loop programme identifies potential stem-loop structures caused by inverted repeat sequences. Examination of K1 ORF 1 and K2 ORF 2 revealed 96 and 76 such structures, respectively, of various sizes and stabilities, on the basis of a minimum 6 bp stem and a maximum loop size of 20 nt. Generally, the longer the stem length and the smaller the loop size (optimally about 6 nt), the greater the stability of the structure. The 'Fold' programme identifies the most stable secondary structures for an RNA molecule. Unfortunately, this programme cannot accommodate sequences longer than about 500 bp, therefore the conformation predicted using this method may not be very representative of the species which actually exist *in vivo*. Nevertheless, analyses of small K2 ORF2 mRNA fragments demonstrated that a high degree of structuring probably existed (see Chapter 7). The 'Terminator' programme searches for prokaryotic factor-independent terminator-like sequences. It does this by locating G+C rich regions of dyad symmetry which lie near poly-U sequences. This led to the identification of 13 regions in the K2 ORF 2 RNA and 9 regions in the K1 ORF 1

RNA with the potential to form terminator-like structures. The three most stable examples are shown in Figure 3.10. Furthermore, the existence of multiple adjacent adenine nt in DNA and RNA sequences has been proposed to be "slippery" to both elongating *E. coli* Rpol and ribosomes (Weiss, cited in Gold & Stormo, 1990). This may lead to slippage of reading frames and premature translational termination. There are extremely A-rich stretches within both ORFs. For example, within the N-terminal 390 bp of K2 ORF 2 there are 3 sequences of 13, 14 and 19 nt, where the adenine-content is 11, 14 and 16, respectively.

Taq-mediated PCR is known to be susceptible, albeit at a low frequency, to the introduction of mismatches (Dunning *et al.*, 1988). PCR-induced artefacts in sequences of this nature could, quite conceivably facilitate the generation of fortuitous termination signals in the correct context for cessation of transcription or translation. There are, unfortunately, relatively few useful restriction sites within the K-plasmid Dpol genes, and this is one of the reasons why PCR was used to clone these sequences. Despite this, all DNA restriction enzyme digestions yielded the expected results, and a limited sequence analysis (~200 nt) of PCR-generated K-plasmid DNA also did not reveal any obvious nt alterations. Although this suggests a low PCR error rate, such have been encountered in other K-plasmid sequences generated by PCR (Soond, Pers. Comm.). The best strategy for avoiding these PCR induced errors, during the generation of clonable fragments, is to use a proof-reading polymerase (eg. Vent polymerase), to replace as much of the PCR generated DNA as possible with the native gene, and to sequence the PCR fragment prior to cloning (eg. using the N.E.B. CirumVent kit).

The *in vitro* transcription experiments demonstrated that despite the highly processive nature of T7 Rpol, transcription of the putative K2 Dpol gene results in the synthesis of multiple aberrant transcripts, the largest being only 1.4 kb long. A similar occurrence was reported by Romanos & Boyd (1988) during attempts to over-express K1 ORFs 2 and 4 in yeast. Furthermore such truncated transcripts are probably more prone to turnover than full length mRNAs, and thus would probably have a very short functional half-life. Any small, prematurely terminated peptides derived from these transcripts would also probably be rapidly degraded.

Figure 3.10 The analysis of the putative K-plasmid DNA polymerase mRNAs for the location of potential prokaryotic rho-independent terminators. Three of the most stable structures for each of the ORFs are shown below. Nine such structures were found for K1 ORF 1 using the 'Terminator' programme in the UWGCG software package. Thirteen were identified for K2 ORF 2.

K2 ORF 2

1. Position in gene= nt 1726-1755
Stem size=7, loop size=4

```
A   A
U   A
 U-A
 G-C
 U-C
 G-U
 A-U
 A-U
 U-A
UAACU G-C UUUAU
```

2. Position in gene = nt 2000-2034
Stem size=8, loop size=5

```
    A
U     G
G     U
  U-A
  A-U
  G-C
  A-U
  C-U
  G-U
  U-A
  A-U
UAGUA U-A UCCAAGU
```

3. Position in gene= 1894-1928
Stem size=8, loop size=4

```
C     A
G     G
  U-A
  A-U
  U-G
  A-U
  U-A
  G-G
  A-U
  U-A
AAAAU A-U CUUUAUAU
```

K1 ORF1.

1. Position in gene= 270-301
Stem size=8, loop size=2

```
A     A
  G-C
  G-C
  U-A
  U-A
  U-G
  U-A
  G-C
UAU G-C UUUUAUUUU
```

2. Position in gene= 1604-1624
Stem size=4, loop size=6

```
  U U
A   U
U   C
  U-A
  G-C
  A-U
UA U-A AAAAU
```

3. Position in gene= 2663-2692
Stem size=5, loop size=3

```
    A
A   G
  U-A
  A-U
  U-A
  A-U
GG C-G GACACUGAUUUUAU
```

At the translational level, a further consequence of the K-plasmids' high A+T content is that the codon bias exhibited by these genes is significantly different from that of *E. coli*. Despite the fact that codon preference, as a factor contributing to efficient gene expression, has generally been ignored, evidence is now accumulating that suggests a lack of cognate charged tRNAs is a major limitation in foreign gene expression. For maximal expression, the codon usage of the heterologous gene must correspond to that of the host. The occurrence of a few unfavourable codons through out a gene probably has little effect upon translation. However, clustering, or a high content of unfavourable codons or high levels of mRNAs containing these codons, could dramatically reduce the maximal level of protein synthesis. The low abundance of some iso-accepting tRNA pools and their subsequent depletion, increases the probability of mis-incorporations, frame-shifting, and possibly ribosome stalling, thus exposing downstream regions of mRNA to RNases. Codon replacement studies in the *S. cerevisiae PGK* gene demonstrated that substitution with synonymous (but less favourable) codons dramatically reduced the level of gene expression (Hoekema *et al.*, 1987). Replacement of 39% of the codons led to a 10-fold reduction in protein levels. Although steady state mRNA levels were also reduced (3-fold), this effect was less significant and suggested that reduced efficiency of expression was due to impaired translational efficiency. More recent experiments with the mouse IgK gene have also demonstrated that optimisation of the codon content (replacement of 115 of the 215 codons) resulted in a 50-fold increase in the steady state product level in *S. cerevisiae* (Kotula & Curtis, 1991).

A comparison of the codon preferences (in terms of the Relative Synonymous Codon Usage) for the K-plasmids, *E. coli*, *S. cerevisiae* and *S. aureus* is shown in Table 3.1. This demonstrates that the expression of K-plasmid genes in *E. coli* would be limited by those codons exhibiting a high RSCU in K-plasmid genes but a low RSCU in *E. coli* (ie. common in K-plasmid genes, but rare in *E. coli*) The most limiting tRNAs would, therefore, be those for isoleucine (AUA), arginine (AGA), glycine (GGA) and leucine (UUA) where the excess requirement by the K-plasmid genes is 250, 266, 102 and 36-fold, respectively. The distribution of these four unfavourable codons in all the K-plasmid genes is summarised in Table 3.3. Despite the differences in codon preference, it should be remembered that the K-

Table 3.3. The Distribution of the Rarest *E. coli* Codons in the K-plasmid ORFs. The % figure represents the proportion of each particular residue that is constituted by the codons. AUA, AGA, GGA and UUA are present in a 250, 266, 102 and 36-fold excess, respectively, in the K-plasmid genes.

Plasmid	ORF	Size (amino acids)	AUA (isoleu)	AGA (Arg)	GGA (Gly)	UUA (Leu)
K1	1	995	59/92 (64%)	31/35 (89%)	20/39 (51%)	42/70 (60%)
	2	1146	32/78 (41%)	24/26 (92%)	30/86 (35%)	53/79 (67%)
	3	428	23/35 (65%)	8/11 (73%)	4/10 (40%)	29/42 (69%)
	4	249	15/23 (65%)	12/13 (92%)	8/16 (50%)	13/22 (59%)
K2	1	224	10/21 (48%)	3/5 (60%)	3/7 (43%)	11/15 (73%)
	2	994	90/111 (81%)	26/28 (93%)	23/36 (64%)	51/79 (65%)
	3	594	42/60 (70%)	23/26 (88%)	15/18 (83%)	30/42 (71%)
	4	597	48/71 (68%)	20/23 (87%)	11/26 (42%)	34/52 (65%)
	5	158	2/14 (14%)	1/1 (100%)	0/5 (0%)	7/12 (58%)
	6	982	104/123 (85%)	38/43 (88%)	35/55 (64%)	45/78 (58%)
	7	132	7/15 (47%)	6/7 (86%)	0/2 (0%)	9/11 (82%)
	8	95	7/10 (70%)	1/2 (50%)	1/3 (33%)	5/11 (45%)
	9	453	41/52 (79%)	15/15 (100%)	11/16 (69%)	29/34 (85%)
	10	103	5/11 (45%)	3/3 (100%)	0/1 (0%)	4/7 (57%)
	11	101	4/11 (36%)	2/3 (67%)	2/3 (66%)	6/8 (75%)

plasmids are maintained and therefore expressed in a variety of other yeast species which also have RSCUs which are significantly different from that of the K-plasmids. It is likely that the levels of K-plasmid gene expression are low, even in *K. lactis*, but problems are compounded when trying to express these proteins to a high level. This is supported by the fact that low levels of some K-plasmid genes have been expressed in heterologous systems (see Table 3.4). Interestingly, Tokunaga *et al.* (1989) reported the successful expression of ORF 4 in *E. coli* as a fusion product with Staphylococcal Protein A, from plasmid pRIT2T. The 58 kDa product was affinity purified using IgG sepharose and used to raise polyclonal antibodies in rabbits. No data, however, was presented concerning the level of recombinant fusion protein production; presumably expression levels were low. Nevertheless, the resulting antibodies specifically recognised both the native and recombinant ORF 4 products.

The detailed analysis of a large number of *E. coli* and *S. cerevisiae* protein coding sequences led to the definition of a species specific combination of the least used codons (Zang *et al.*, 1991). Low usage codons were clearly avoided in genes encoding abundant proteins, and cognate tRNA abundance was roughly proportional to codon usage. Not surprisingly, several of the rarest codons in *E. coli* are common in K-plasmid genes: the two rarest, AGA and AUA, are common in both K1 and K2. Sharp & Li (1986) defined a rare *E. coli* codon as that having a RSCU of 0.05 or less. There are ten of these triplets. Examination of the codon composition of K1 and K2 Dpol genes reveals that the total rare codon content, calculated on this basis, is approximately 13% and 16%, respectively. The majority of these rare codons are, however, made up of the two rarest triplets alone (Table 3.3 and 3.5), AGA (arginine) and AUA (isoleucine). A very high proportion of the total number of arginine and isoleucine residues, are specified by the rare codons. For example, 89% and 93% of the arginine residues in the K1 and K2 Dpols, respectively, are encoded by AGA (the second rarest *E. coli* triplet in highly expressed genes). Furthermore, the single rarest codon in *E. coli* (AUA) specifies 64% and 81% of the isoleucine residues in K1 ORF 1 and K2 ORF 2, respectively. This distribution of rare codons is also non-random, and this is illustrated by Table 3.5, where the small gene fragments all have rare codon contents above the average value for the whole gene. The definition of rare codons as those occurring with a frequency of less than 10% of each amino acid in the highly expressed ribosomal protein genes (de Boer

Table 3.4. The Expression of K-plasmid Genes in Heterologous Systems. A summary.

Plasmid	ORF	Putative function	Details	Reference
K1	3	immunity	Expressed in yeast from a nuclear vector, but not visible by SDS-PAGE. Also expressed in *E. coli* as protein A fusion from the vector pRIT2T; affinity purified & antibodies raised.	Tokunaga *et al.*, 1987
	4	toxin γ-sub unit	Expressed in Yeast, but not visible by SDS-PAGE. Detected by killing phenotype.	Tokunaga *et al.*, 1989
K2	1	?	Reportedly expressed in *E.coli*, although no evidence presented to support this.	McNeel & Tamanoi, 1991
	2	Dpol	Part of gene resynthesized, expressed as GST fusion in *E.coli*	Ambrose (this thesis)
	5	?	Reportedly expressed in *E.coli*, but no evidence presented to support this. Also expressed in *K. lactis* from a recombinant K1 derivative. Not visible by SDS-PAGE, but detected *via* integral epitope tag.	McNeel & Tamanoi, 1991; Schaffrath (Pers. Comm.).
	6	Rpol	Small fragments expressed in*E.coli* as GST fusions, but products generally insoluble. Also expressed in *E. coli* as fusion with β-galactosidase. Antibodies raised to both GST & β-gal fusions, but did not recognise anything in yeast extracts.	Brophy (Pers.Comm); Fleming & Meacock, (unpublished results)
	7	Rpol	Expressed in *E.coli.* as insoluble GST fusion. Antibodies raised to the fusion did not recognise anything in yeast extracts. Also expressed, in *E. coli* using pT7-7, in native form, and with C-terminal epitope & affinity tags.	Brophy (Pers.Comm), Schaffrath (Pers.Comm.).
	10	ds DNA binding protein	Expressed in *E.coli* , detected by gel retardation. Also produced as MS2 Rpol fusion in *E. coli*. Antibodies raised to both, & used purify native protein from yeast extracts.	Tommasino, 1991, McNeel & Tamanoi, 1991

Table 3.5. Codon Usage Comparison of Two Over-expressing and Two Non-expressing K-plasmid Gene Fragments as Fusions with GST in *E. coli*. Both the Rpol fragments are from K2 ORF 6, the Dpol fragments from K2 ORF 2. Induction of the Rpol recombinants led to the synthesis of high levels of insoluble protein, but induction of the Dpol recombinants led to simultaneous loss of non-fusion carier and lack of fusion product. Induction of the Dpol recombinant plasmids also gave rise to a much reduced growth rate, when compared to the Rpol recombinants (Brophy, Pers.Comm.). The percentage figures in brackets represent the proportion of each amino acid specified by the rare triplets.

Amino acid	Rpol (760 bp) 75% A+T	Rpol (270 bp) 73.5% A+T	Dpol (390 bp) 78% A+T	Dpol (470 bp) 74% A+T
Ala	3	2	5	3
Arg	12 (100%)	6 (100%)	5 (100%)	5 (80%)
Asn	20	3	8	10
Asp	18	7	9	13
Cys	7	2	0	4
Gln	2	0	2	3
Glu	17	5	13	12
Gly	7 (43%)	4 (75%)	1 (0%)	7 (43%)
His	5	2	2	3
Iso	36 (78%)	3 (100%)	15 (93%)	21 (95%)
Leu	20 (60%)	12 (75%)	9 (44%)	11 (55%)
Lys	30	9	24	17
Met	8	2	3	3
Phe	8	2	7	5
Pro	9	2	0	3
Ser	18	5	11	7
Thr	9	5	8	5
Trp	4	0	0	3
Tyr	14	6	7	15
Val	8	6	1	8
Proportion of residues specified by rare codons	22%	23%	18%	21%

Table 3.6 The distribution of clustering (defined as two or more immediately adjacent) of single rare *E. coli* codons in the K-plasmids Dpol genes. There are also many mixed doublets and some mixed triplets of rare codons. In K2 the glycine rich sequence GGA-GGA-GGA-TAT-GGA at residues 774-778 may be particularly problematical.

Doublet	Location in K1 ORF 1 (residues)	Location in K2 ORF 2 (residues)
Arg-Arg (AGA-AGA)	340-1, 976-7	690-1
Iso-Iso (ATA-ATA)	389-90, 954-5	414-5, 443-4, 460-1, 470-1, 493-4, 714-5, 836-7, 873-4, 915-6
Leu-Leu (UUA-UUA)		567-8, 647-8
Gly-Gly (GGA-GGA)		271-2, 774-6

& Kastelein, 1990) raises the number of rare triplets to 23. When applied to the K1 and K2 Dpol genes, this increases the proportion of rare codons to 34% and 35%, respectively. These figures illustrate the highly unfavourable codon bias of these genes.
Surprisingly, none of the rarest codons in *S. cerevisiae* are common in either K1 or K2. As *S. cerevisiae* also has a higher genomic A+T content (approximately 60%) than *E. coli* (approximately 49%), and the K-plasmids are stably maintained in *S. cerevisiae*, there may be a greater chance of successful K-plasmid gene expression in *S. cerevisiae*.

The occurrence of immediately adjacent rare codons has also been proposed to have extremely deleterious effects upon gene expression. The proposals made by Varenne & Lazdunski (1986) were confirmed by observations that the presence of adjacent minor arginine codons (AGG or AGA) resulted in 50% +1 ribosomal frame-shifting (Spanjaard & Van Duin, 1988). This, they proposed, was due to minor tRNA depletion. Interestingly, both K1 and K2 Dpol genes contain 'AGA doublets', at residues 341-2 in K1, and 690-1 in K2 (Table 3.6). Examination of the +1 reading frame confirms that, if such slippage occurred in the K-plasmid Dpol genes, non sense codons would very quickly be encountered; in fact after 16 codons in K1 ORF1 and 22 in K2 ORF 2. Analysis of DNA sequence data reveals that the K-plasmid genes (most noticeably K1 ORF 1, K2 ORF 2 and K2 ORF 6) contain multiple doublets of the four rarest codons. The most common of these is the isoleucine AUA-AUA sequence, this being an even rarer triplet in *E. coli* than the minor arginine codons.

Gene fusions have advantages in stabilisation, purification and compartmentalisation. However, fusion proteins often have reduced solubility and may form inclusion bodies. This is not a problem if the protein is required simply for its immunogenic properties. The simplest fusion is the addition of a secretion signal which if correctly processed yields the native N-terminal (recently review by Stader & Silhavy, 1990). Fusion of the recombinant peptide to the C-terminus of a control protein not only proteolytically stabilises these products, but also has the advantage that the expression signals for the first gene are not disrupted, facilitating efficient transcriptional & translational initiation and early elongation. The optimisation of these processes often allows the synthesis of the parental non-fusion moiety

to high levels. The coupling of foreign genes to these sequences would therefore, under normal circumstances, be expected to result in high level synthesis of the fusion product. If the product is soluble, affinity "handles" can be used to facilitate isolation. The *Staphylococcal* protein A system is one such case (reviewed by Nilsson & Abrahmsen, 1990). Dual affinity fusions, with the recombinant protein sandwiched between two heterologous domains, each with affinity for a different ligand, allows full length product recovery using two purification steps, as used for the expression of the Ad Dpol (Rekosh *et al.*, 1985). This is especially suitable for proteins particularly susceptible to proteolysis. Several affinity systems have also been designed based on small synthetic peptides (see Chapter 5).

The pGEX vectors direct high level synthesis of GST. The conjugation of K-plasmid Dpol gene fragments to this peptide, however, appears to result in the complete destabilisation of the product. Neither the expected recombinant fusion or the non-fusion peptides are seen upon induction of these strains. This is not due to their synthesis in the form of insoluble aggregates. Even small (390 bp) fragments of the K2 Dpol gene, which Northern analysis has demonstrated are fully transcribed by T7 Rpol, do not allow low level product synthesis. This also does not seem to be due to frame shifting, or out-of-frame translational fusions, as the non-fusion GST would probably be seen. Nonsense codons occur at a high frequency in both the +1 and +2 reading frames. For example, at the extreme N-terminus of K1 ORF 1, stop codons occur in the +1 and +2 reading frames after 8 and 6 codons, respectively. Similarly, in K2 ORF 2 they occur in the +1 and +2 reading frames 5 and 4 codons from the N-terminus. This was confirmed by the inversion of the internal 470 bp *Bgl* II-*Sau* 3a fragment of K2 ORF 2 in pGEX1 (pCA14). A nonsense codon is encountered only 7 triplets into the K-plasmid sequence. This resulted in the cessation of translation and high levels of the parental GST moiety were observed (Figure 3.8).

Similar problems were experienced during attempts to express the K2-encoded Rpol (Brophy & Meacock, unpublished results). Full length ORF 6 could not be over-expressed in *E. coli*, but some small fragments of ORFs 6 and 7 were expressed as GST fusions, although these products were mainly insoluble. Analysis of the A+T content and codon bias (in terms of the total number of rare codons and of each individual rare codon) of these small ORF 6 and

ORF 7 fragments failed to reveal any significant differences from those exhibited by the small K2 Dpol fragments (Table 3.5). The ORF 7 gene has, however, recently been expressed in a soluble form, at low levels, in *E. coli* using the T7 Rpol system (Schaffrath, Pers.Comm.), but attempts to express K2 ORF 5 with the same system proved unsuccessful. Similarly, recent attempts to express the K2 ORF 4 product as a GST fusion, also failed (Soond, Pers.Comm.).

The possible reasons for these failures are numerous and complex. For example the *E. coli* Rpol may be unable to transcribe even short K-plasmid sequences, due to the presence of strong pause sites or terminator-like sequences within the template. Inefficient transcription or translation can, for different reasons, lead to destabilisation of the fusion mRNA or peptide. The ability of *E. coli* to express several small fragments of ORF 6 as fusions with GST (Brophy & Meacock, unpublished results) suggests that perhaps this failure to express has a less obvious reason. For example, product toxicity or a DNA binding activity. The PCR-cloned N-terminal 390 bp K2 ORF 2 fragment would be expected to carry a strong positive charge at physiological pH, whereas the internal 470 bp fragment of the same gene would have no net charge. At the N-terminus of K2 ORF 2, within the space of 49 amino acids (see Figure 4.1), there are no less than 16 lysine residues (ie. 32.7%). At physiological pH, this fragment of the protein would carry a stronger net positive charge than the 390 bp fragment. Is this a DNA binding domain?

Among the few reports of successful K-plasmid gene expression in heterologous systems are those concerning the toxin and immunity factors. Both these gene products do have properties which are relatively easy to assay, so low level expression could be detected, even if the products were not visible by SDS-PAGE. The insertion of a K1 fragment containing ORFs 3 and 4 (the immunity and toxin γ-sub unit, respectively) into an *E. coli-K. lactis* shuttle vector led to the low level expression of ORF 3 (Tokunaga *et al.*, 1987). This was detected by the immunity of recombinant plasmid-bearing strains to the K-plasmid toxin, although the level of immunity observed was significantly lower than that obtained with the native system, implying that nuclear transcription of this gene may be less efficient than cytoplasmic transcription. Expression of the immunity factor was, however, only detected

in the presence of K2, suggesting a requirement for a K2 encoded factor. Furthermore, although ORF 4 was also present on this recombinant plasmid, no mention was made as to whether this was expressed. The effects of ORF 4 expression could have been masked by the immunity phenotype. Other experiments, however, suggest that ORF 4 was probably not expressed (Tokunaga *et al.*, 1989; Romanos & Boyd, 1988). Northern analyses of transcripts produced from K1 ORFs in yeast nuclear vectors demonstrated that in all cases, except (significantly) ORF 3, no native transcripts were detected. These results support the view that K-plasmid genes are not preceded by recognised yeast nuclear promoters, are not efficiently transcribed by yeast nuclear or *E. coli* Rpols and that expression of ORF 3 in such a vector was purely fortuitous.

Following on from these experiments, attempts to express K1 ORF 4 from its native promoter in a yeast nuclear circular vector, proved unsuccessful. Replacement with the constitutive *PGK* promoter resulted in unusually low transformation frequencies of competent *S. cerevisiae* (Tokunaga *et al.*, 1989). Further replacement of the promoter, with the regulated *GAL7* promoter led to the restoration of normal transformation frequencies, and subsequent induction led to rapid arrest of culture growth, a state which could be prevented by the co-expression of the immunity gene.

There have also recently been two further reports of K-plasmid gene expression in *E. coli*, both concerning ORF 10 of K2 (Tommasino, 1991; McNeel & Tamanoi, 1991). ORF 10 appears to encode a small (13 kDa) hydrophilic protein with a high basic amino acid content (about 27% lysine). Tommasino (1991) expressed this ORF as a fusion with MS2 Rpol, which was clearly visible by SDS-PAGE analysis of crude soluble protein extracts. Polyclonal antibodies raised to the product demonstrated the presence of the ORF 10 protein in crude extracts prepared from K-plasmid bearing *K. lactis* strains. Cell fractionation and co-immunoprecipitation studies subsequently demonstrated that this protein was associated with the DNA of both k-plasmids in the same sub-cellular (cytoplasmic) fraction.

McNeel & Tamanoi (1991) further defined the DNA binding property of the ORF 10 product. Gel retardation experiments, in conjunction with deletion analysis, demonstrated that the terminal regions of K1 and K2 were specifically recognised and bound by a protein

present in cytoplasmic extracts of K-plasmid bearing yeast strains. The identity of the terminal region recognition factor was confirmed by the expression of the three smallest ORFs of K2 (ORFs 1, 5 and 10), in *E. coli*. These expressed proteins, however, were not visible by SDS-PAGE of crude cell extracts and only extracts prepared from the ORF 10-expressing strain catalysed the characteristic gel shift. The failure of the other extracts to produce a gel shift, however, could also reflect a lack of expression. Deletion analysis and DNase I foot-printing subsequently defined the regions involved in binding as bp 107-183 in K1 and 126-179 in K2. As the ITRs of K1 and K2 exhibit no sequence homology, the ORF 10 product (TRF1) presumably recognises a structural feature within these regions (similar to p6 of ø29). TRF1 was then partially purified from yeast extracts using ds DNA affinity chromatography (in the same manner as used previously to purify adenoviral replicatory proteins), and when immobilised on nitrocellulose filters, it specifically bound [^{32}P]-labelled K-plasmid ITR fragments.

The over-expression of the small genes or small fragments of the larger K-plasmid genes in *E. coli* or *S. cerevisiae* is possible (Table 3.4), but success depends on a number of factors. The most important of these are probably whether the gene is efficiently transcribed and translated by the host. Small gene fragments which are not prematurely terminated and encode relatively hydrophilic products, with a minimal (absolute) content of rare codons are most likely to be expressed. The nature of the gene product may also limit the level of expression. Toxicity of a heterologous gene product can usually be diagnosed by difficulty in obtaining transformants with constitutive expression vectors (eg as with pRIT5) or by reduced growth rate with induced regulated systems (eg. with the pET and pGEX systems). This may have contributed to the problems experienced during attempts to express the Dpol genes. The absolute quantities of the K1 ORF 3 and 4 (Tokunaga *et al.*, 1987 & 1989), and K2 ORF 1, 5 and 10 products (McNeel & Tamanoi, 1991; Schaffrath, Pers. Comm.) were small and were not detectable by SDS-PAGE of crude cell extracts. Furthermore, in each case a specific assay (eg. toxicity, immunity, gel retardation or antibody binding) was available to aid identification of the product.

CHAPTER 4

EXPRESSION OF A SYNTHETIC PORTION OF THE PUTATIVE K2 DNA POLYMERASE GENE IN *E. COLI*.

4.1 INTRODUCTION.

Experimental evidence discussed in Chapter 3 suggests that the difficulties experienced during attempts to express the putative Dpol genes of K1 and K2, is due to the inability of *E. coli* to transcribe and translate this DNA efficiently. This has been illustrated by the synthesis of truncated transcripts, arising as a result of premature transcriptional termination, and by the fact that short sections of the genes which have been shown to be completely transcribed, do not yield any obvious protein product. Both of these problems have probably arisen as a consequence of the high A+T content of the K-plasmid genome.

If these genes had an average A+T content, and a correspondingly different codon bias, expression of these genes would probably be relatively straightforward. This proposal could be tested by the resynthesis of these genes. Such technology was, in fact, used during the early heterologous gene expression experiments in *E. coli*, with the small peptides somatostatin (Itakura *et al.*, 1977) and insulin (Goeddel *et al.*, 1979). Nowadays, although the isolation and cloning of heterologous genes is commonplace, chemical gene synthesis has assumed a new importance, because of the difficulty experienced during attempts to over-express certain proteins in *E. coli* (see Chapter 3). Many of these problems have been attributed to discrepancies between the codon preferences exhibited by the heterologous gene and the host expression system. This is probably one of the reasons why the native K-plasmid Dpol genes were not over-expressed in *E. coli*. Resynthesis of a gene not only allows the codon content to be optimised with regard to the host, but also allows the engineering of restriction enzyme sites within the gene, to facilitate construction and subsequent manipulation. Some examples of genes which were chemically resynthesized and subsequently over-expressed to high levels in *E. coli* are summarised in Table 4.1.

To date, the most common method of constructing synthetic genes has remained the synthesis of multiple oligonucleotides of about 20 bases, corresponding to both strands. These are simply annealed, ligated and cloned. However, as the technology for DNA synthesis has improved and become more accurate, reliable and reproducible, the length of synthetic oligonucleotides which can be efficiently synthesised has increased (Table 4.1).

Table 4.1 Selected examples of the chemical synthesis and expression of heterologous genes in *E. coli*.

Protein.	**Size.**	**Oligos.**	**References.**
Somatostatin	15 amino acids	8, of 7-16 nt	Itakura *et al.*, 1977
Insulin A & B chains	21 & 30 amino acid	A chain: 12, of 10-14 nt; B chain: 18, of 10-12 nt	Goeddel *et al.*, 1979 Crea *et al.*, 1978
α-IFN	517 bp	67, of 10-20 nt	Edge *et al.*, 1981
Ribonuclease S	318 bp	66, of 10-22 nt	Nambiar *et al.*, 1984
Bovine intestinal Ca^{2+} binding protein	76 residues, 297 bp	34, of upto 24 nt	Brodin *et al.*, 1986
Sperm whale myoglobin	507 bp	23, of 45-48 nt	Springer & Sligar, 1987
Bovine pancreatic ribonuclease A	363 bp	Not disclosed	Nambiar *et al.*, 1987
Horseradish peroxidase	308 bp	40, of 32-60 nt	Jayaramen *et al.*, 1991
Horse heart myoglobin	153 residues (523 bp)	12, of 60-103 nt	Guillemette *et al.*, 1991

Genes can now be synthesised from fewer, but longer oligonucleotides. For example, whereas 67 different oligonucleotides, of 10-20 nt were used in the synthesis of the 517 bp α-interferon (Edge *et al.*, 1981), only 12 oligonucleotides, of 60-103 nt were used during the synthesis of the 523 bp horse heart myoglobin (Guillemette *et al.*, 1991). Obviously, the oligonucleotides have to be designed so as to minimise the possibility of fortuitous non-specific annealing and any subsequent sequence re-arrangements or deletions. For this reason, sequencing of both strands is also essential. The advent of PCR technology has also allowed the amplification of the final full length product (by using the end two oligonucleotides as primers), the use of smaller quantities of oligonucleotides and product sequencing prior to cloning. This approach was used during the construction of the horseradish peroxidase gene (Jayaramen *et al.*, 1991).

The large size of both the K-plasmid Dpol genes, however, precludes their complete resynthesis, on economic grounds. Therefore, on the basis of previous work, a small fragment of the extreme 5' region of K2 ORF 2 was resynthesised, the most important criteria in its redesign being to incorporate the optimal codon bias exhibited by highly expressed *E. coli* genes.

4.2 RESULTS.

4.2.1 Construction of the Synthetic Gene.

Most gene construction has been based upon the synthesis of overlapping oligonucleotides corresponding to both DNA strands, which are subsequently annealed and ligated. Only a very short section of K2 ORF 2 was to be resynthesised, and in order to simplify construction and minimise cost, the design of the oligonucleotides was such that the maximum length of ds DNA could be obtained using the minimum number and length of oligonucleotides. This was achieved by restricting the length of the complementary sequence in the two oligonucleotides to 18 nt, so that each effectively binds only the terminal region of the other and acts as a primer for a Klenow-type filling-in mechanism.

On this basis, four oligonucleotides of 97, 98, 100 and 107 nt were designed to facilitate the reconstruction of the extreme 330 bp (109 amino acid) 5' section of ORF 2 of K2 in two stages, each of the products being independently cloned and sequenced, and

subsequently spliced together (Figure 4.2). The oligonucleotides were designed such that the gene fragment had a *Bam*H I site at the 5' and 3' termini, a central *Eco*R I site and a 3' *Bgl* II restriction site. These sites were incorporated to allow initial cloning into a pIC vector, where the two DNA fragments could be assembled and sequenced, and then subsequently sub-cloned into pGEX2T and pET3a, to allow expression in the form of a fusion and non-fusion protein. In fact, due to time limitations, only the first section of this gene (amino acids 1-54), using oligonucleotides CA17 and CA18, was constructed but the design for the entire fragment in shown in Figure 4.1 and 4.2.

The resynthesised gene fragment incorporated the optimal codon bias exhibited by highly expressed *E. coli* genes. The high lysine content of this fragment resulted in a several long stretches of A nucleotides (the preferred lysine codon being AAA) which might prove problematic during annealing and cloning (Mallinder, Pers.Comm.) and could also possibly lead to depletion of the cognate tRNAs. Therefore, regions of adjacent lysine triplets were disrupted by the substitution of the preferred codon by the next most favoured (AAG) lysine triplet (Figure 4.1), as is seen in the native K2 ORF 2.

Oligonucleotides CA17 and CA18 were purified from a 20% polyacrylamide gel using UV-shadowing to detect their position. They were then annealed and extended using a Klenow-type elongation reaction. The ds DNA fragment was then digested with *Eco*R I and *Bam*H I, purified by PAGE and cloned between the *Bam*H I-*Eco*R I sites of pIC20H. Colonies bearing the recombinant plasmid were identified by insertional inactivation of the *lacZ* product. Restriction analysis subsequently confirmed the presence of the DNA fragment. This plasmid was named pCA28. Both strands of the DNA fragment were then sequenced using the forward and reverse universal primers. No deviations from the expected sequence were observed (Figure 4.3).

4.2.2 Expression of the Synthetic Gene.

4.2.2a. As a Non-fusion Protein.

Gene 10 of the bacteriophage T7 contains particularly efficient translational initiation signals and these have been incorporated into the pET3 vectors. Part of the reason for the previous lack of expression of the K2 ORF 2 fragments in the pET3a derivatives pCA9

Figure 4.1. The redesigned amino acid and nucleotide sequence of the N-terminal portion of K2 ORF2 incorporating an internal *Eco*R I site, *Bam*H I sites at each end, a *Bgl* II site at the 3' end and the optimal codon bias for expression in *E. coli*. The top line (*) also shows the native K2 ORF 2 nucleotide sequence. The second half of this fragment downstream of the central *Eco*R I site, was not synthesised due to time constraints. The table at the foot of the page shows the codon bias of the native (old) and resynthesised (new) gene fragment.

(*) ATG AAT GAT GAA TTA GCA TTC TTA AAT TCT CAA ATA GAT GAC TAT TCC GAT GAT ATA GAA TTT CTA
S M N D E L A F L N S Q I D D Y S D D I E F L
Ser met asn asp glu leu ala phe leu asn ser gln iso asp asp tyr ser asp asp iso glu phe leu
TCC ATG AAC GAC GAA CTG GCT TTC CTG AAC TCT CAG ATC GAC GAC TAC TCTGAC GAC ATC GAG TTC CTG

AGA ACA GAA GTA TTT AGA ATA AAT AGA TAT TCT AAA TCT ATA TAT TCA TTA ATG AAA AAT AGT TCT AAG
R T E V F R I N R Y S K S I Y S L M K N S S K
arg thr glu val phe arg iso asn arg tyr ser lys ser iso tyr ser leu met lys asn ser ser lys
CGT ACC GAA GTT TTC CGT ATC AAC CGT TAC TCT AAA TCT ATC TAC TCT CTG ATG AAA AAC TCT TCT AAA

GAA GCA ATA AAA GAA ACT ATA GAA TTT AAT AAA TAT GGT CAA GAA ATA CAT TCT AAA ATA AAA GAA AAA
E A I K E T I E F N K Y G Q E I H S K I K E K
glu ala iso lys glu thr iso glu phe asn lys tyr gly gln glu iso his ser lys iso lys glu lys
GAA GCT ATC AAA GAA ACC ATC GAA TTC AAC AAA TAC GGT CAG GAA ATC CAC TCT AAA ATC AAA GAA AAA
*Eco*R I

GAA ATA AGT CTT ATA GAA GCA ATA AAG CTA ATG GAA AAG AAA TAT AAA TAT TTA AGA AGT CTT AAT AAA
E I S L I E A I K L M E K K Y K Y L R S L N K
glu iso ser leu iso glu ala iso lys leu met glu lys lys tyr lys tyr leu arg ser leu asn lys
GAA ATC TCT CTGATC GAA GCT ATC AAA CTG ATG GAA AAG AAA TAC AAA TAC CTG CGT TCT CTG AAC AAA

ACT AAA GAC AAG TCT AAA AAG AAG AAA ATT CAC ACA AAA GAT ATA GAA ACT TTG
T K D K S K K K K I H T K D I E T L
thr lys asp lys ser lys lys lys lys iso his thr lys asp iso glu thr eu Stop
ACC AAA GAC AAA TCT AAA AAG AAA AAG ATC CAC ACC AAA GAC ATC GAA ACC CTG TAA AGA TCT GGA TCC
Bgl II *Bam*H I

nt	U			C			A			G			nt
	αα	old	new	αα	old	new	αα	old	new	αα	old	new	
U	F	3	0	S	6	11	Y	6	0	C	0	0	U
U	F	1	4	S	1	0	Y	0	6	C	0	0	C
U	L	4	0	S	1	0	stop	0	0	stop	0	0	A
U	L	1	0	S	0	0	stop	0	0	W	0	0	G
C	L	2	0	P	0	0	H	1	0	R	0	4	U
C	L	0	0	P	0	0	H	1	2	R	0	0	C
C	L	2	0	P	0	0	Q	2	0	R	0	0	A
C	L	0	9	P	0	0	Q	0	2	R	0	0	G
A	I	1	0	T	3	0	N	6	0	S	3	0	U
A	I	0	13	T	0	5	N	0	0	S	0	0	C
A	I	12	0	T	2	0	K	14	16	R	4	0	A
A	M	2	2	T	0	0	K	6	4	R	0	0	G
G	V	0	1	A	0	3	D	5	0	G	0	0	U
G	V	0	0	A	0	0	D	2	7	G	0	0	C
G	V	1	0	A	3	0	E	12	11	G	1	1	A
G	V	0	0	A	0	0	E	0	1	G	0	0	G

Figure 4.2 The sequences and arrangement of the oligonucleotides designed for reconstruction of the N-terminus of K2 ORF 2.

Oligo1 (98mer).
(5')
CCCCCGGATCCATGAACGACGAACTGGCTTTCCTGAACTCTCAGATCGACGACTA
*Bam*HI

CTCTGACGACATCGAGTTCCTGCGTACCGAAGTTTTCCGTATC (3')
annealing region

Oligo3 (97mer).
(3')
GGCTTCAAAAGGCATAGTTGGCAATGAGATTTAGATAGATGAGAGACTACTTTT
annealing region

GAGAAGATTTCTTCGATAGTTTCTTTGGTAGCTTAAGCCCCC (5')
*Eco*RI

Oligo 2 (110mer)
(5')
CCCCCGAATTCAACAAATACGGTCAGGAAATCCACTCTAAAATCAAAGAAAAAGA
*Eco*RI

AATCTCTCTGATCGAAGCTATCAAACTGATGGAAAAGAAATACAAATACCTG (3')
annealing region

Oligo4 (107mer)
(3')
CTTTATGTTTATGGACGCAAGAGACTTGTTTTGGTTTCTGTTTAGATTTTTCTT
annealing region

TCTAGGTGTGGTTTCTGTAGCTTTGGGACATTTCTAGACCTAGGCCCCC (5')
*Bgl*II *Bam*HI

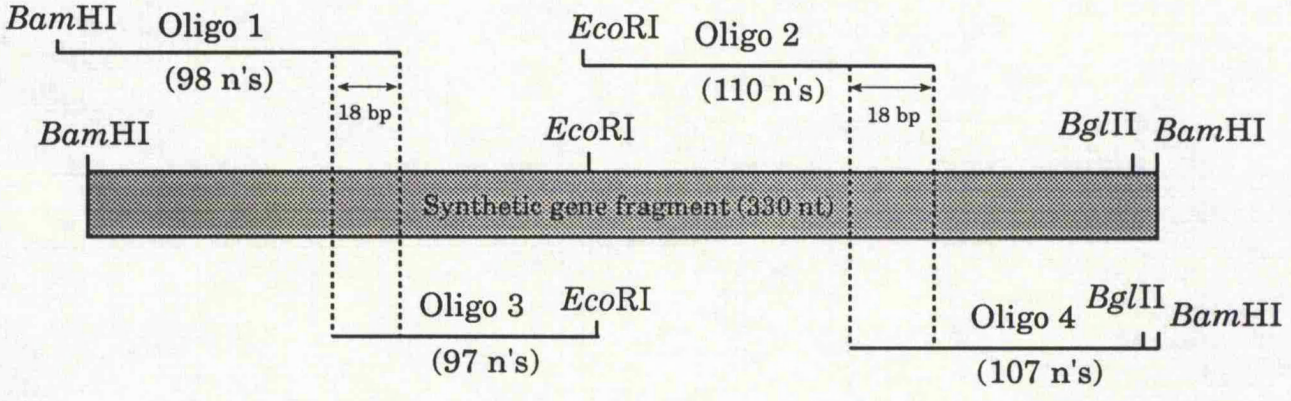
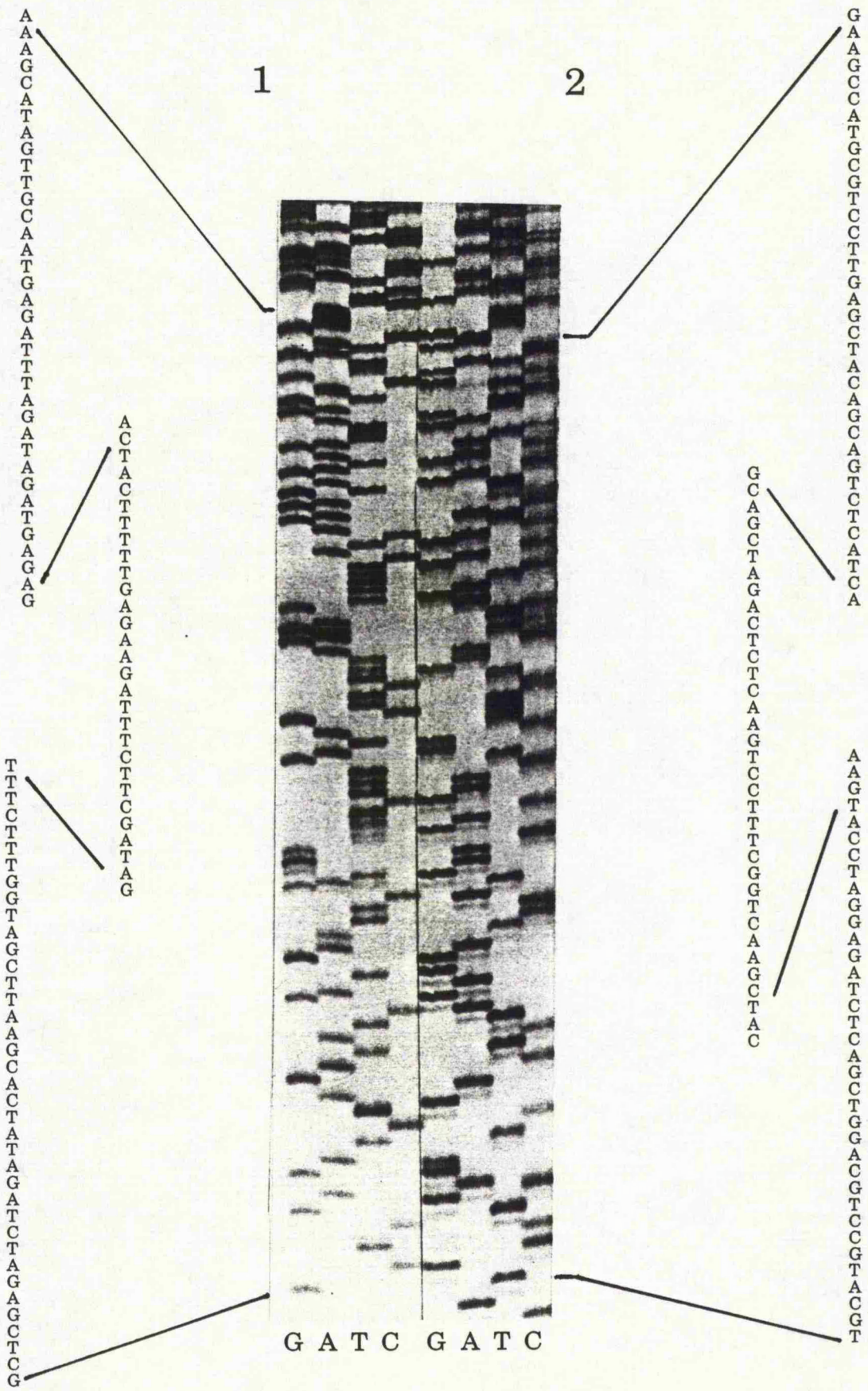


Figure 4.3 The nucleotide sequence of the resynthesised fragment of K2 ORF 2, incorporating an optimal *E. coli* codon bias. Sequence 1 was obtained using the -20 universal primer, sequence 2 with the reverse universal primer.



and pCA13, may have been the deletion of the s10 sequence in pCA8. Although this removed the duplicate *Nhe* I and *Eco*R V sites, it also deleted the first eleven codons of the gene 10 product; the extreme 5' coding region of the gene may influence the efficiency of translational initiation, as has been reported for some *E. coli* genes. Therefore, the 168 bp resynthesized K2 ORF 2 fragment was isolated from pCA28 as a *Bam*H I-*Bgl* II fragment and was sub-cloned into the *Bam*H I site of pET3a itself, producing an in-frame translational fusion. The structure of the recombinant plasmid (pCA29), including the presence of the synthetic gene fragment, was confirmed by restriction analysis. Although there was no control over the orientation of the fragment in recombinant plasmids, this could be distinguished using the *Eco*R I sites within the synthetic gene and pET3a (Figure 3.4).

Transformation of competent *E. coli* BL21(DE3) with pCA29, and subsequent induction and analysis of protein fractions from multiple independently isolated clones, yet again failed to reveal any sign of the expected 5 kDa recombinant protein. The [^{35}S]-methionine radiolabelling of proteins synthesised *in vivo*, using rifampicin inhibition of host transcription also yielded negative results, as did linked *in vitro* transcription-translation studies using *E. coli* cell free extracts. However, *in vitro* transcription from plasmid pCA29 demonstrated that high levels of the appropriate mRNA were being synthesised by the T7 Rpol (figure 4.4). This suggests that either this RNA was unstable *in vivo*, or that the resultant protein was rapidly degraded *in vivo*. Similar problems have been encountered on numerous occasions by researchers trying to over-express small proteins. Such problems can usually be circumvented by synthesis as fusion proteins.

4.2.2b. As a Fusion Product.

The synthetic gene fragment was isolated from pCA28 as an *Bam*H I-*Eco*R I fragment and cloned into the *Bam*H I-*Eco*R I sites of pGEX2T, producing an in-frame translational fusion with the GST moiety. The presence of the DNA fragment in recombinant plasmid pCA30 was confirmed by restriction analysis. IPTG induction of *E. coli* bearing this plasmid led to the synthesis of high levels of recombinant protein, to at least 10% of total cell protein, as revealed by SDS-PAGE analysis of crude cell extracts (Figure 4.5). The

Figure 4.4 Analysis of the plasmid pCA29. **Panel A.** 0.8% w/v agarose gel showing restriction enzyme digestions of plasmids prepared from *E. coli* transformants during the initial pCA29 construction. Lanes 6-10, *Eco*R I digest of five clones (1-5) of pET3a containing the K2 ORF 2 *Bam*H I-*Bgl* II synthetic gene fragment in both orientations. The correct orientation is that which yields the largest *Eco*R I fragment (ie. lanes 6, 7 & 10). Lane 5, *Eco*R I-*Bam*H I digestion of pET3a. Lanes 2-4, *Bam*H I digestions of clones 1, 4 & 5. **Panel B.** *In vitro* transcription from clones 1, 4 & 5, electrophored on a 1.5% w/v agarose gel. Lane 9 is RNA size markers (kb). The expected size of the transcript of the synthetic gene, including the upstream & downstream sequences from pET3a, was approximately 0.38 kb. **Panel C.** Autoradiograph from the Northern blot of the gel in Panel B, probed with a radiolabelled 5' fragment of K2 ORF 2.

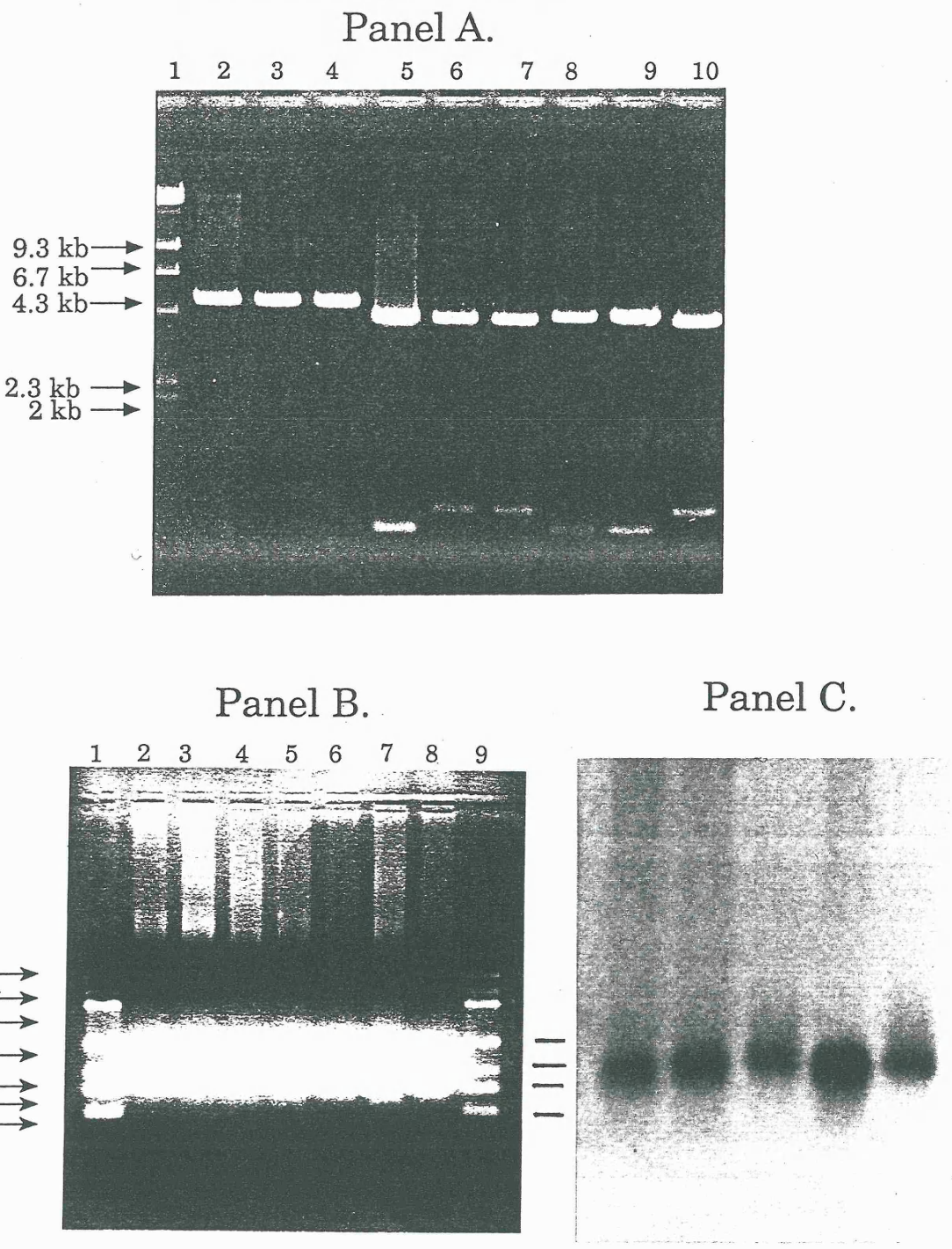
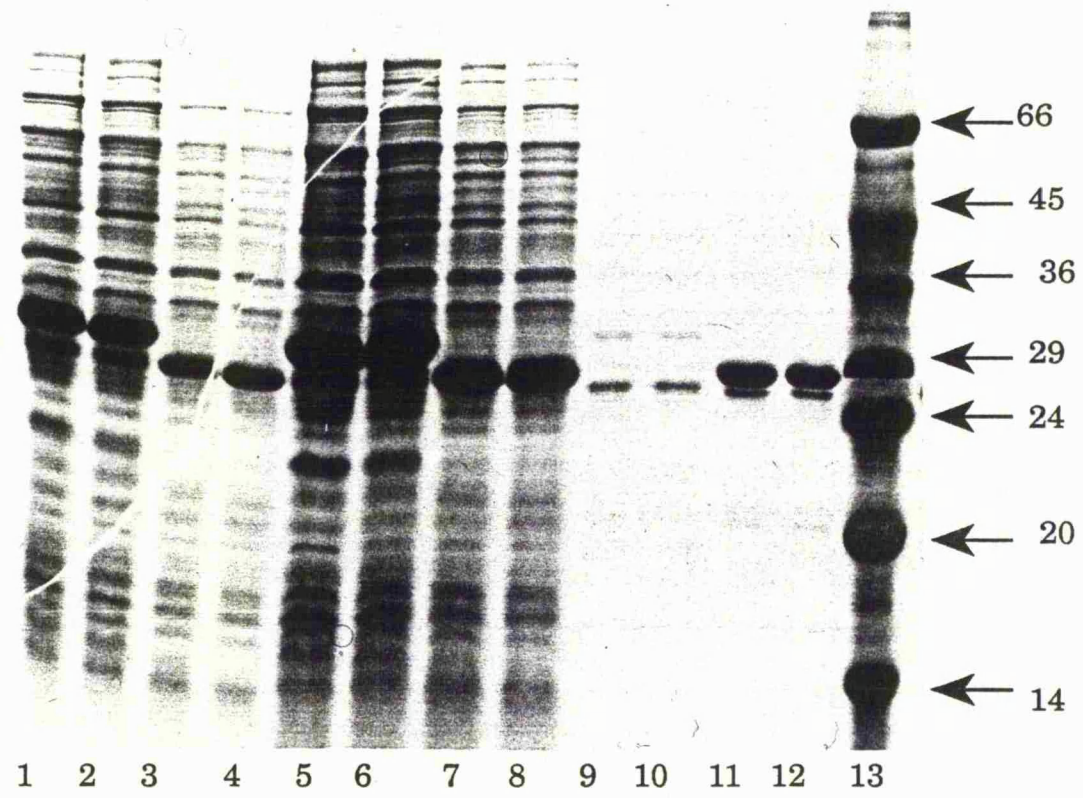
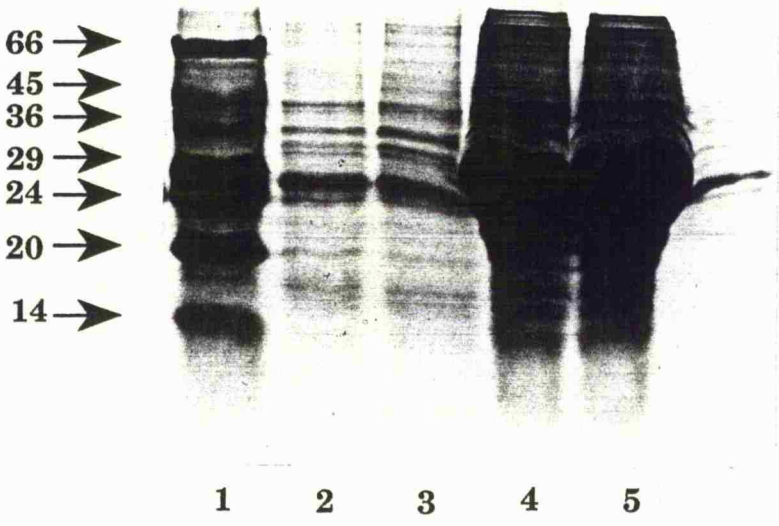


Figure 4.5 Panel A. 15% SDS-PAGE of the synthetic gene and the SP2 epitope of K2 ORF 2 (see Chapter 5), as C-terminal fusions with GST. Lanes 1-8 show total protein samples of IPTG induced cultures; lanes 9-12 show protein isolated by means of the glutathione agarose mini-batch method, from IPTG induced cultures. Duplicate samples were prepared from separate bacterial clones bearing the same recombinant plasmid, and two different loadings were made for each sample. Lanes 1, 2 , 5, 6, 9 & 10 are pCA30; lanes 3, 4, 7, 8, 11 & 12 are pCA31; lane 13 is molecular weight markers. **Panel B.** 15% SDS-PAGE of the most insoluble protein fractions from duplicate IPTG induced cultures of pCA31 (lanes 2 & 3) and pCA30 (lanes 4 & 5). Lane 1 is molecular weight markers.





parental GST protein (27.5 kDa) exhibited a shift in electrophoretic mobility consistent with the addition of an extra 54 amino acids (5 kDa), the recombinant fusion protein having a molecular weight of approximately 32 kDa. Attempts to purify this recombinant protein using affinity chromatography revealed that only extremely small amounts of the protein could be isolated, despite the fact that the control non-fusion GST molecule was easily purified using this method (Figure 4.5). This suggested that either the addition of the extra 54 amino acids destabilised the protein during the isolation procedure, despite the presence of protease inhibitors in the cell lysis buffer, or that the product had a reduced solubility. Subsequent analysis of cellular protein fractions revealed that the fusion product was insoluble (Figure 4.5).

4.3 DISCUSSION.

The design and construction of a synthetic gene allows the incorporation of an optimal codon bias for the host in which it is to be expressed. Furthermore, careful design also allows problems at the transcriptional level to be avoided. Should the use of the optimal codon bias result in the generation of sequences which may be problematic transcriptionally, the offending codons can be replaced with the next most favourable triplet, although whether the substitution of certain codons can be accommodated depends upon how faithful the researcher wishes to remain to the original protein.

In most cases, attempts to express native K-plasmid genes in *E. coli* have been unsuccessful. Although the primary reason for this appears to be transcriptional, the *E. coli* Rpol being unable to retain processivity upon these templates, much of the problem can also be attributed to the low efficiency translation of available messengers. The extreme N-terminal region of the K2 Dpol was, therefore resynthesized, the coding sequence containing a near optimal bias for those codons found in the highly expressed *E. coli* genes. Although only the first 53 amino acid fragment was actually resynthesised, examination of the complete projected 109 residue segment reveals some interesting features. In order to maximise translational efficiency, 69 of the 109 codons (ie. 63%) needed to be altered. This required 79 nt changes (24% of the coding sequence). However, 66 of these 79 nt substitutions (83.5%) were located in the third position of the

codon, and resulted in 59 of the 66 codon alterations. Of the 109 amino acids encoded, 20 specified lysine residues; 16 of these were located within a stretch of 49 amino acids in the second half of the peptide. Thus, although the A+T content of highly expressed *E. coli* genes is only about 49%, that of the projected synthetic gene fragment remained high, 61.8%. This is, however, a significant reduction from the 78.5% of the native sequence. If the clustering of the lysine residues is taken into account, the A+T content of the first half of the K2 ORF 2 gene fragment is about 58%, but that of the second half approximately 65.6% (Figure 4.1). Therefore, although the nt content of a gene is affected by its origin (ie. in this case being cytoplasmic), it is also affected by the nature of the product it encodes. In this case, the high lysine content (the preferred codon being AAA) contributes significantly to the high A+T.

The failure of the synthetic gene fragment to be expressed in pET3a, despite containing an optimal codon bias and a near optimal nucleotide content, may be due to a number of possible causes. The small (5 kDa) product may have been rapidly degraded by the host. The problem of degradation of heterologous proteins in *E. coli* is well recognised, and small peptides are particularly prone to rapid degradation. This instability was observed during the early heterologous gene expression experiments with somatostatin (Itakura *et al.*, 1977), insulin (Goeddel *et al.*, 1979) and β-endorphin (Shine *et al.*, 1980). In these cases the problem was circumvented by the synthesis of the peptides as C-terminal fusions with β-galactosidase. As the recombinant protein expressed from the pET3a vector was only about 65 amino acids long (including the 11 residues of the T7 gene 10 product) such degradation was possible. However, the radiolabelling of proteins synthesised *in vitro*, using cell free linked transcription-translation systems, is usually very sensitive and control experiments using pET3a recombinant plasmids confirmed this (Figure 3.3). The pET3a-*E.coli* BL21 (DE3) system itself is not completely trouble-free. The variability in the BL21 (DE3) host, in terms of its ability to over-express heterologous proteins has already been noted (Section 3.2.1.2c). Recent analysis of this system, using a series of 86 in-frame deletions of the HSV ribonucleotide reductase, revealed that the nature of the sequence immediately downstream of the initiating ATG can drastically effect the level of expression (Furlong *et al.*, 1992). Only about 40% of the resultant clones actually expressed

detectable levels of the protein. The basis of this effect remains unclear, but it does emphasise the need for detailed experimentation with each gene being expressed.

The insolubility of the GST-synthetic gene product is intriguing; the addition of such a small (5 kDa) peptide to a relatively large (27 kDa) carrier being sufficient to dramatically reduce solubility. The over-expression of many recombinant genes in *E. coli*, both prokaryotic and eukaryotic, is often limited by the production of insoluble cytoplasmic or periplasmic protein aggregates (reviewed by Kohno *et al.*, 1990). The exact cause(s) of this insolubility generally remain to be elucidated, but there have been a number of attempts to identify such 'solubility' factors. In several cases, these results have subsequently been used, by extrapolation of primary sequence data, to predict recombinant protein solubility. However, factors relating to solubility are probably more closely related to secondary, than primary, structural features. Nevertheless, two, the 'charge average' and the 'turn forming residue fraction' were identified by Wilkinson & Harrison (1991), although based upon the data they presented, these correlations seem, at best, to be tenuous. The charge average is the net charge on the peptide, at neutral pH, divided by the total number of residues. The charge average of the synthetic gene fragment is not easy to determine, the calculation requiring complex computer analyses, but a simplified version, based upon the numbers of charged and uncharged residues at physiological pH, expressed as a fraction of the total fragment size as can be used. Those residues which are charged at neutral pH are assigned positive values, whereas those uncharged (ie. aliphatic and aromatic amino acids) are given negative values, and the difference between these values is then averaged over the length of the peptide. This reveals that the redesigned 54 amino acid N-terminal fragment has a 'charge average' of approximately -0.3. As the proportion of charged residues increases, as it would if, for example, the next 56 residues of the native protein were taken into account, this figure tends towards zero, the charge average of the complete 109 amino acid fragment now being -0.06. The higher the final value, the more likely the peptide is to be soluble. It is possible that the fusion of a peptide with such a charge average to the GST molecule could have reduced the solubility of the entire fusion product sufficiently for it to precipitate out of solution.

Turns in a polypeptide chain are the most difficult structures for proteins to form, so a high content of such turn-forming residues may indicate a slow folding protein. However, examination of the resynthesised gene fragment reveals that the 'turn forming residue fraction' (primarily proline residues) is very low (in fact, zero). Therefore, this factor is unlikely to be the cause of the insolubility. Alternatively the yield of soluble protein may be governed by the competing reactions for non-specific aggregation and folding under conditions of drastically increased concentration. At high local concentrations of unfolded or partially folded peptide, arising due to the insufficient time for correct folding of the newly synthesised peptides, aggregation predominates leading to inclusion body formation and precipitation (Kiefhaber *et al.*, 1991). The most straightforward way of avoiding this problem, by allowing the folding reactions to predominate over those for aggregation, may, therefore, be by reducing the rate of heterologous protein synthesis. This has been accomplished simply by reducing the growth temperature (Schein, 1990) or the induction rate (Blum *et al.*, 1992). This approach has also recently been successful with the pET vectors (Furlong *et al.*, 1992).

The lack of compartmentalisation in *E. coli*, with peptides being placed in an inappropriate reducing cytosolic environment which prevents the interactions necessary for correct folding, may also contribute to product insolubility. In eukaryotic cells compartmentalisation (with different redox and pH environments), protein interactions (foldases, unfoldases, chaperonins) and post-translational modifications (phosphorylation, hydroxylation, glycosylation, proteolysis) all significantly affect protein solubility. There may, however, be ways to mimic these conditions in *E. coli*, for example by secretion or by the co-expression of 'folding helper' molecules (eg. heat shock proteins). Such an effect has been reported for the HSP 70 chaperonin (Blum *et al.*, 1992).

Although most of the GST-Dpol fusion product was insoluble, affinity chromatography did allow the purification of very small amounts of the soluble fusion protein, as can just be discerned in Figure 4.5 (Panel A, lanes 9 and 10). However, attempts to increase the proportion of the GST-Dpol fusion product in a soluble form, were not made. Despite this, the large amounts of the fusion product mean that it can easily be purified. Solublisation,

by using strong chaotropic reagents (eg. 6M urea), would then allow subsequent removal of the GST carrier moiety, via the Thrombin cleavage site at the junction of the GST-Dpol peptides. So much protein is now available that the generation of polyclonal antibodies, should be fairly routine.

CHAPTER 5.

MANIPULATION OF THE K-PLASMIDS: THE ESTABLISHMENT OF RECOMBINANT K-PLASMIDS BY HOMOLOGOUS RECOMBINATION *IN VIVO* AND THEIR USE IN ISOLATION OF K-PLASMID GENE PRODUCTS.

5.1 INTRODUCTION.

Since the elucidation of the nt sequences of K1 and K2 there have been numerous attempts to over-express the putative K-plasmid genes in *E. coli*, *S. cerevisiae* and *K. lactis*. Whilst such attempts have, generally, been unsuccessful (Stark *et al.*, 1984; Stam *et al.*, 1986; Tokunaga *et al.*, 1987; Romanos & Boyd, 1988; Wilson & Meacock, 1988; Brophy, Pers. Comm.; Soond, Pers. Comm.; Ambrose, This Thesis), there have been some notable exceptions (Table 3.4). However, much of the experimental data suggests that the expression (at both the transcriptional and translational levels) of K-plasmid genes in *E. coli* or *S. cerevisiae* and *K. lactis* nuclear vectors is very inefficient. As an alternative it may be possible to engineer the K-plasmids themselves to facilitate identification and isolation of their own proteins.

Until relatively recently, the utility of this approach has been limited by the lack of technology for handling these plasmids. Like other 'linear plasmids', K1 and K2 require intact TPs, specific DNA sequences within the ITRs, and self encoded protein factors for their replication. Any attempts to manipulate these plasmids must, therefore, take these considerations into account. In addition, the K-plasmids appear to replicate in the cytoplasm. Removal of the TPs, circularisation and retransformation of yeast leads to localisation of the plasmids to the nucleus (Stam *et al.*, 1986; Tokunaga *et al.*, 1987). This results in inactivation of the plasmid-borne genes. These, it appears, cannot be recognised or transcribed by yeast nuclear Rpols (Romanos & Boyd, 1988). The crucial factor in manipulating these elements is, therefore, the ability to re-introduce the plasmids, with their intact TPs, back into the cytoplasm. This maybe possible by electroporation in the near future, but preliminary attempts in our laboratory have been unsuccessful (Reay, Pers. Comm.).

A strategy for manipulating these plasmids has, however, now been devised and developed in several laboratories, including our own. This involves modification of K-plasmid DNA sequences in *E. coli*, the transformation of K-plasmid bearing *K. lactis* strains with linear K-plasmid-derived DNA fragments and their subsequent homologous recombination *in vivo*, into the native K-plasmids. The introduction of heterologous sequences by this means, however, tends to be restricted to those portions of the plasmids which are known

to be non-essential for replication and maintenance. Both K1 and K2 have very little intergenic DNA, ORFs occupying more than 95% of the available DNA, and most of the remainder is accounted for by the terminal sequences which are probably essential for replication. Insertion must, invariably, be into a structural gene. The majority of ORFs on K2 have yet to be assigned functions, but 3 of the ORFs on K1 are known to be dispensable in terms of replication and maintenance; those encoding the toxin and immunity (ORFs 2, 3 and 4). These are, therefore, the best target sites for integration, despite the fact that the toxin/immunity phenotype could be valuable self-selection markers. However, these could eventually be regained by incorporation into a single hybrid K1-2 linear cytoplasmic plasmid, following the definition of the essential and subsequent removal of redundant genes.

5.1.2 Manipulation of the K-plasmids *In Vivo*.

The first attempt to modify the K-plasmids *in situ* was made by the insertion of a selectable marker, the *S. cerevisiae URA3* gene (including its endogenous promoter) flanked by about 1 kb of both K1 terminal sequences, into *K. lactis*, to allow complementation of a nuclear *K. lactis URAA* mutation (Louvencourt *et al.*, 1983). In all of the prototrophs examined the marker had, however, become localised to the nucleus as a circular plasmid. Similarly, the insertion of the *S. cerevisiae LEU2* gene into K2 also resulted in all of the prototrophs which were isolated and examined, containing circularised nuclear derivatives of this plasmid (Fujimura *et al.*, 1987). Autonomous replication of these plasmids was probably due to the fortuitous occurrence of K-plasmid DNA sequences with the ability to exhibit ARS activity.

The first truly recombinant cytoplasmic K-plasmid was generated by the insertion of a kanamycin resistance gene (Km^R), into K1 ORF 2 (Tanguy-Rougeau *et al.*, 1990). When the Km^R gene promoter was used to direct expression of the selectable marker, no transformants were obtained. Southern analysis demonstrated that the gene had integrated at the expected site, but Northern analysis revealed that the gene was not transcribed. Replacement of the endogenous promoter with that of K1 ORF 1 resulted in a high level resistance to kanamycin. Northern analysis subsequently demonstrated the presence of the

expected 1.3 kb transcript. Similar results were obtained using the *S. cerevisiae LEU2* gene (Kamper *et al.*, 1989a, b & 1991). When the *LEU2* gene was inserted into K1 with its homologous promoter, the recombinant plasmids became localised to the nucleus. Some of these nuclear plasmids were circular. Others, however, retained their linearity, but were stabilised by the specific addition of yeast telomeric sequences. The replacement of the *LEU2* transcriptional (promoter and terminator) signals with those of K1 ORF 2 subsequently resulted in the cytoplasmic localisation and expression of the selectable marker. The integration of both the Km^R and *LEU2* genes resulted in the generation of recombinant K1 plasmids of different size to the native K1. The gross structure of both the linear cytoplasmic derivatives was, nevertheless, similar to native K1, containing both TP and ITRs. Furthermore, both recombinant K1 plasmids could be segregated from native K1, in the presence of K2, by extensive sub-culturing under selective conditions and the resultant colonies were non-killers, but retained the immunity phenotype.

5.1.3 The Use of Epitope and Affinity Tags to Isolate Proteins.

Fusion proteins have a number of advantages in recombinant protein production. These do, however, have limitations in their subsequent use for antibody production, especially if the conjugant moiety cannot be removed after purification. The stabilising peptide is often relatively large and highly antigenic (eg. *E. coli* β-galactosidase). Therefore only a small proportion of the antibodies raised may be specific for the required peptide. This is reminiscent of the problems encountered with coupling synthetic peptides to carrier molecules (Chapter 6).

One of the most useful properties of fusion proteins is the ability to specifically isolate the products from crude samples, by means of the affinity 'handles' constituted by the stabilising moiety. This is exemplified by the affinity chromatography used to isolate protein A and GST derivatives. Another drawback to using fusion proteins concerns the limited resemblance these chimeric products bear to the native protein. If the size of the stabilising moiety can be minimised, whilst retaining the affinity characteristics, this could maximise the resemblance of the product to the native protein and facilitate the generation of antibodies which specifically recognise the non-fusion protein in cell extracts. The

rapid advances in recombinant DNA technology have, therefore, allowed the development of small peptide 'tags', which when fused to the recombinant protein allow specific detection and isolation, *via* an affinity matrix.

In this study two such tags, an 'epitope' and a chelating-peptide tag have been used. The former corresponds to a continuous epitope (Chapter 6) and basically consists of a short antigenic synthetic peptide. When incorporated into a recombinant protein, this sequence is specifically bound by antipeptide antibodies. This can be exploited, for product isolation, by the construction of an anti-peptide antibody affinity matrix.

Transition metal cations have long been known to interact with certain amino acids and proteins. This observation has been exploited in the development of immobilised metal ion affinity chromatography (IMIAC). At least two metal chelating matrices, based on a derivatised hydrophilic support matrix (Sepharose or Sephadex), have been developed for this purpose, yielding matrices with different metal chelating properties. Some metals are of greater significance than others due to their ability to interact strongly with certain amino acids. For example, Ni^{2+} exhibits a particularly strong affinity for histidine residues. Small histidine polymers have been used successfully, in conjunction with Ni^{2+} IMIAC, in several cases to isolate recombinant proteins. A poly-histidine chelating peptide, containing 4-6 histidine residues fused to either the N- or C-terminus, has been used to isolate mouse dihydrofolate reductase from crude *E. coli* extracts (Hochuli *et al.*, 1988). The HIV-1 reverse transcriptase was isolated by the incorporation of a histidine heptapeptide into either the N- or C-termini, from over-expressing *E. coli* strains (Le Grice *et al.*, 1990). The human serum response factor was also isolated from recombinant vaccinia virus using an N-terminal hepta-histidine tag (Janknecht *et al.*, 1991). A similar system, utilising a poly-arginine peptide, has also been reported (Sassenfeld & Brewer, 1984).

Having developed the technology to manipulate the K-plasmids *in situ* and express heterologous genes from cytoplasmic derivatives, this system can now be utilised to analyse the K-plasmid genes themselves. If the genes can be modified, for example by the attachment of affinity handles, and re-introduced into the plasmids in a functional form, this technique may facilitate isolation of the encoded proteins.

5.2.1 RESULTS.

5.2.1 Construction of the Recombinant pGKL1 DNA polymerase.

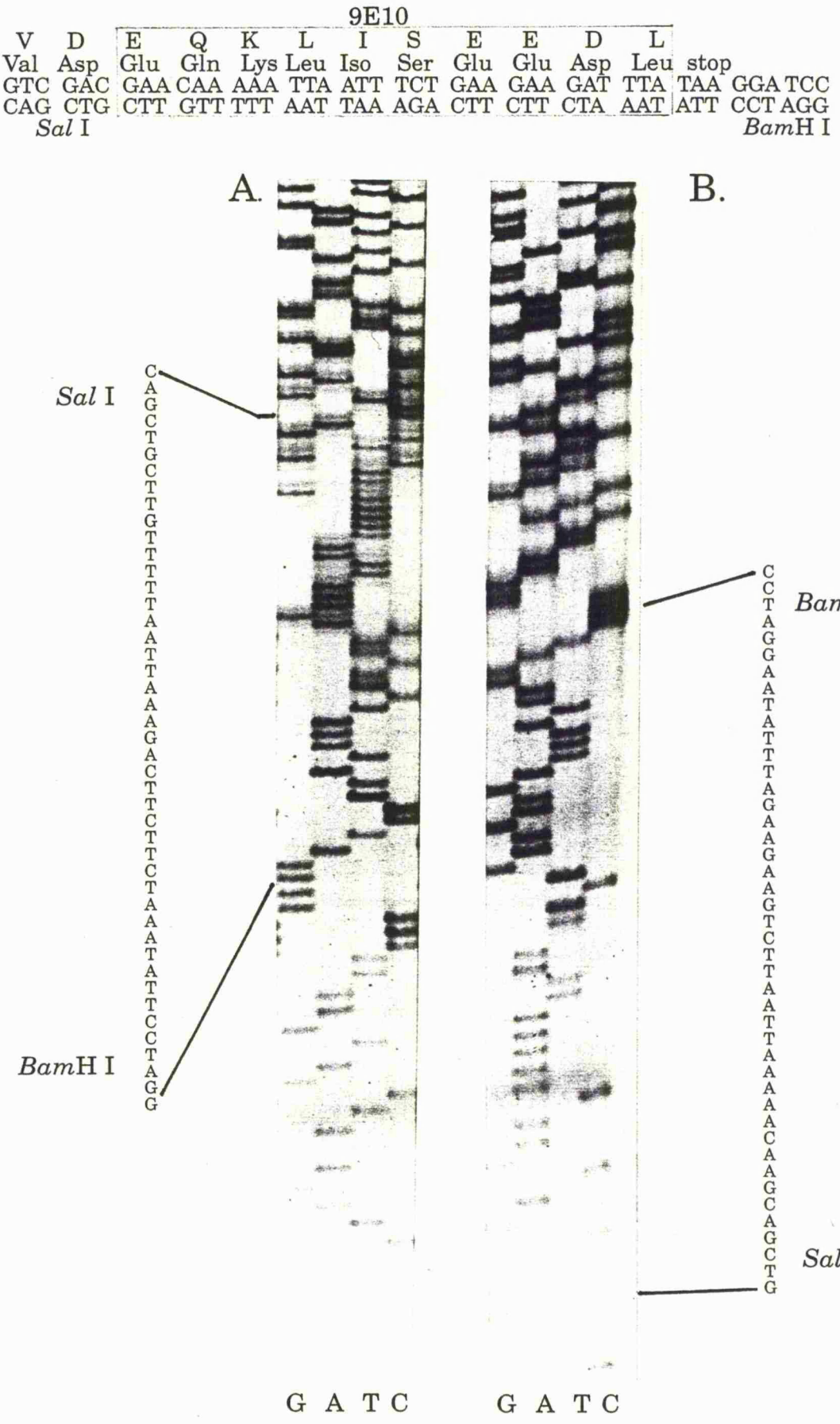
5.2.1a Construction of the epitope tag.

Six monoclonal antibodies specific for the human *c-myc* proteins were raised to two different synthetic peptides (Evan *et al.*, 1985). Although all six monoclonals recognised the immunoblotted *c-myc* protein, only five immunoprecipitated the native protein. The other monoclonal, denoted 9E10, only recognised electroblotted *c-myc*, but was the most sensitive probe for use in Western analyses. The sequence recognised by the 9E10 monoclonal (EQKLISEEDL) has subsequently been fused to and used to detect human insulin (Dr. C. Hadfield, Pers. Comm.) and human T-cell receptor over-produced by *S. cerevisiae* (Dr. C. Hewitt, Pers. Comm.).

The two short oligonucleotides CA13 and CA14, each of 39 nt, corresponding to the peptide sequence recognised by the 9E10 monoclonal antibody were synthesised (Figure 5.1). They were designed such that the 9E10 recognition sequence could be fused, in frame, onto the 3' end of the K1 Dpol gene. Furthermore, upon annealing, the resultant small ds DNA fragment had *Sal* I and *Bam*H I half-sites at the 5' and 3' ends, respectively, facilitating direct cloning. The 9E10 fragment also utilised the optimal K-plasmid codon bias.

The oligonucleotide solution was ethanol precipitated, quantified, diluted and annealed, and then cloned into *Bam*H I - *Sal* I cut pIC19H (Marsh *et al.*, 1984), to yield plasmid pCA18. Although insertional inactivation of the *lacZ* gene product was used to select recombinant pIC plasmids, this was found to be unreliable with small DNA fragments. The presence of recombinant plasmids often gave a range of light and dark blue colonies, the in-frame insertion allowing translational 'read through' and retention of some β-galactosidase activity. Transformation frequencies were a better guide to a successful cloning step. The small DNA fragment could also be visualised by PAGE of *Sal* I-*Bam* H I digestions of relatively pure plasmids. Finally, to confirm the presence and structure of the DNA fragment a number of pCA18 clones were sequenced (see Figure 5.1).

Figure 5.1 The nucleotide sequence of the 9E10 epitope tag in the cloning vector pIC19H. The structure of the tag is shown below. **Sequence A** was obtained using the -20 universal primer, and **Sequence B** with the reverse universal primer.



5.2.1b Insertion of the chelating peptide.

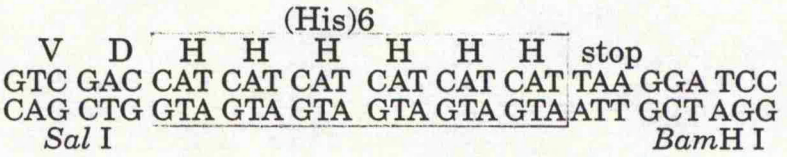
In a similar manner, two oligonucleotides of 27 nt corresponding to a poly-(hepta-) histidine tag were designed and synthesised. This sequence also had terminal *Sal* I (5') and *Bam* H I (3') half sites and incorporated an optimal K-plasmid codon bias. This was cloned into pIC20R, yielding plasmid pPB1 (P. Brophy, Leicester Biocentre). The presence of the tag was also confirmed by DNA sequencing (Figure 5.2).

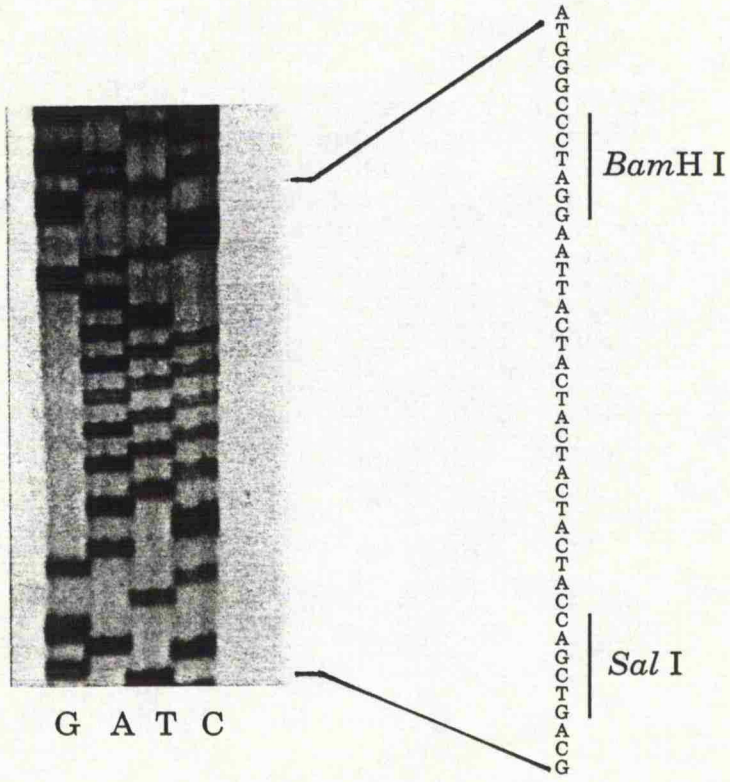
5.2.2 Cloning of the K1 ORF1.

The PCR primers CA1 and CA7, incorporating terminal *Eco*R I (5') and *Sal* I (3') sites, were used to selectively amplify K1 ORF1. The primer CA7, corresponding to the 3' end of ORF 1, was designed to allow an in-frame translational fusion between K1 ORF 1 and the epitope and chelating tags. For this reason, the TAA nonsense codon at the extreme 3' end of the gene was omitted. The 3 kb K1 ORF 1 fragment was selectively amplified, using Vent polymerase to minimise any PCR-induced errors. This was then phenol extracted, ethanol precipitated and treated with Klenow to ensure that all DNA termini were blunt. PCR sometimes results in the non-specific addition of adenine residues to the 3' end of the newly synthesised DNA strands, the presence of which can reduce the efficiency of blunt end cloning (see Chapter 7). The blunt ended PCR product was then gel purified and cloned into the *Sma* I site of pBSM13^{+}, yielding plasmid pCA25. The structure of the 3' end of the PCR amplified K1 ORF 1 was then confirmed by DNA sequencing.

The 1.7 kb fragment of K1 ORF 1 from pCA25 was then excised using a *Pst* I-*Sal* I digestion, and was cloned into plasmid pCA20; the resultant plasmid pCA26, containing an in-frame translational fusion of K1 ORF 1 with the poly-histidine tag. Complete cleavage of the immediately adjacent *Pst* I and *Sal* I sites in both pIC19H and pIC20R was, however, found to be highly inefficient during this cloning step. This appears to have been due, primarily, to the inability of *Pst* I to cut efficiently at terminal sites. In an attempt to circumvent this problem, the 0.6 kb *Pst* I fragment from pEMBL-4K containing the functional Km^{R} gene (C. Hadfield, Leicester Biocentre) was inserted into the *Pst* I of pCA18 and pPB1. Colonies bearing these recombinant plasmids (pCA19 and pCA20,

Figure 5.2 The nucleotide sequence of the poly-histidine tag, in the cloning vector pIC20R. The tag is clearly visible, but the cause of the shadowing on this gel is not known. The sequence of the tag is:





respectively) were selected by their Km^R. Linearisation of pCA19 and pCA20 with *Sal* I, and subsequent cleavage with *Pst* I allowed efficient digestion of both sites, as revealed by the excision of the 0.6 kb Km^R gene.

5.2.3 Construction of the histidine Tag Integrative Cassette.

In order to be able to insert the modified K1 ORF 1 back into native K1 by homologous recombination *in vivo*, and then to segregate the native and recombinant K1, a selectable marker and a K1 sequence downstream of ORF 1 was required. A trp^+ auxotrophic derivative of *K. lactis* MG1/2, was constructed by the insertion, using homologous recombination, of the *S. cerevisiae URA3* gene into the chromosomal *K. lactis TRP1* locus. This *K. lactis* strain (JK1) was designed for use in conjunction with the plasmid pMS11 (Figure 5.3) which contained the *K. lactis TRP1* gene under the control of the K1 ORF 2 expression signals (both constructed and provided by S. Soond, Leicester Biocentre).

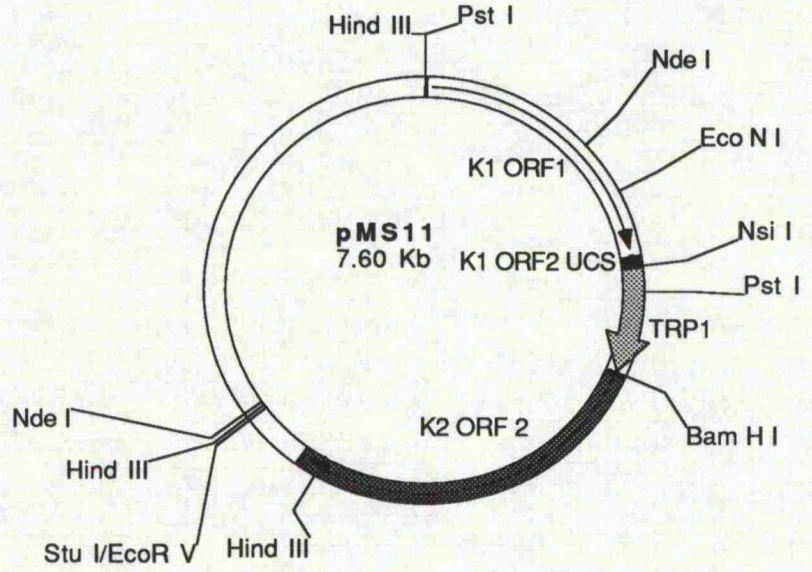
Work at this stage concentrated on the histidine tagged gene; if experiments progressed satisfactorily the construction of the epitope tagged gene was to be completed later. The *TRP1* gene, together with the 3' end of K1 ORF 2, was excised from plasmid pMS11 by a *Nde* I digestion. The cohesive ends of this 4 kb DNA fragment were then filled-in and the fragment was cloned into *Sma* I-cut pCA21. The structure of the resultant recombinant plasmid pCA23, is also shown in Figure 5.3.

5.2.4 Establishment of the recombinant K1 in *K. lactis*.

In order to allow recombination of the recombinant DNA fragment with the target sequence in the endogenous native K1, the recombinant DNA has to be introduced in a linear form. Furthermore, as it is the termini of the linear fragment which are responsible for the recombination, these must correspond to sequences within the native K1. These criteria were accommodated by an *Nde* I-*Hind* III digest of plasmid pCA23. This contains an *Nde* I site approximately 1 kb from the 3' end of ORF 1 and the *Hind* III site lies at the extreme 3' end of ORF 2. Homologous recombination and integration of this fragment (Figure 5.4) was predicted to result in approximately 1.2 kb of the 5' region of K1 ORF 2 being deleted, the insertion of an additional 0.7 kb of the *TRP1* gene, a 1 kb duplication of the 3' section of ORF 1 and the reformation of a complete, but modified, ORF1. The

Figure 5.3 A. Structure of the plasmid pMS11. The *TRP1* gene and the downstream portions of K1 ORF 2 were cut out by restriction with *Nde* I. This fragment was then blunt end ligated into the *Sma* I site of pCA22 to yield pCA23. **B**. pCA23 was then linearised with *Nde* I and *Hin*d III and used to transform *K. lactis* JK1. Recombinant K1 bearing *K. lactis* strains were selected by their TRP prototrophy.

A.



B.

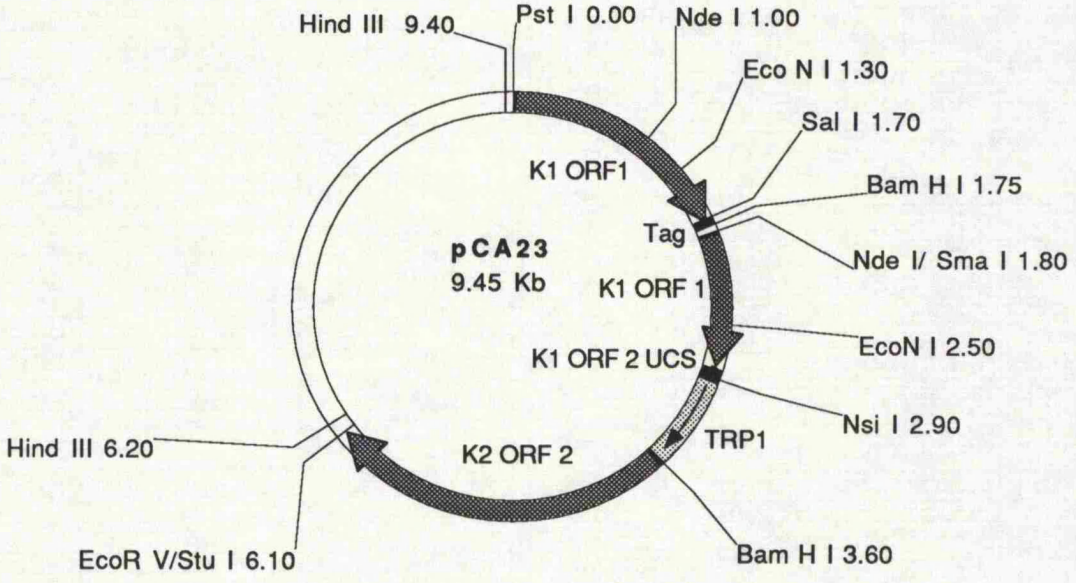
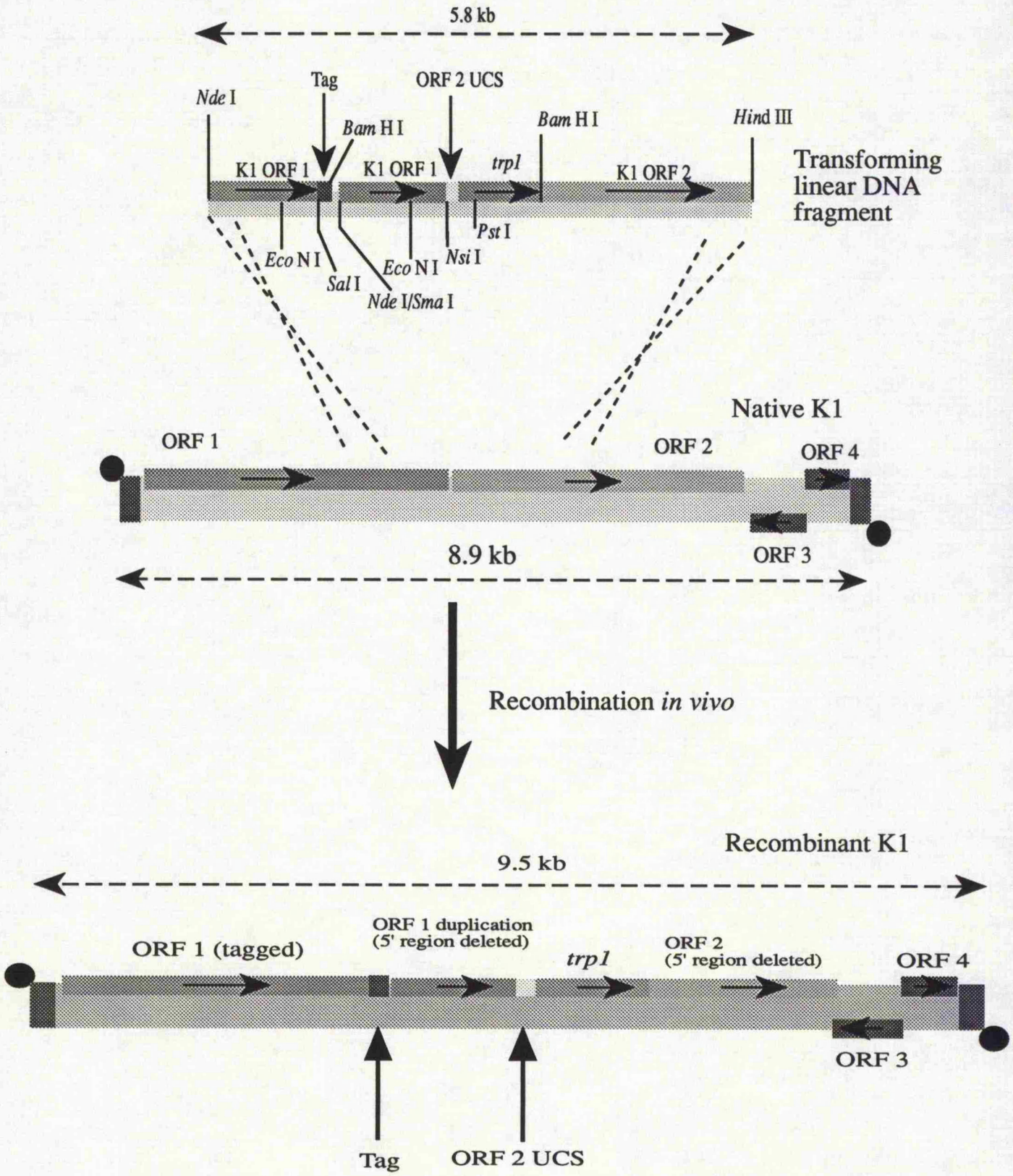


Figure 5.4 The strategy for the re-introduction of a modified version of the endogenous K1 ORF 1 gene. The linear 5.8 kb pCA23-derived *Nde* I-*Hin*d III integrative cassette was used to transform *K. lactis* JK1 bearing both native K1 and K2. The resultant recombinant K1 plasmids were selected by their *trp1* phenotype.



recombinant version of K1 should, therefore, be of approximately 9.4 kb. Aliquots of the *Nde* I-*Hin*d III linearised plasmid digest were then used directly to transform competent *K. lactis* JK1. Transformants (trp^{+}) were generally visible after 7-14 days incubation at 30°C. Analysis of the linear plasmid content of these strains revealed the presence of additional K1 derived plasmids (Figure 5.5A).

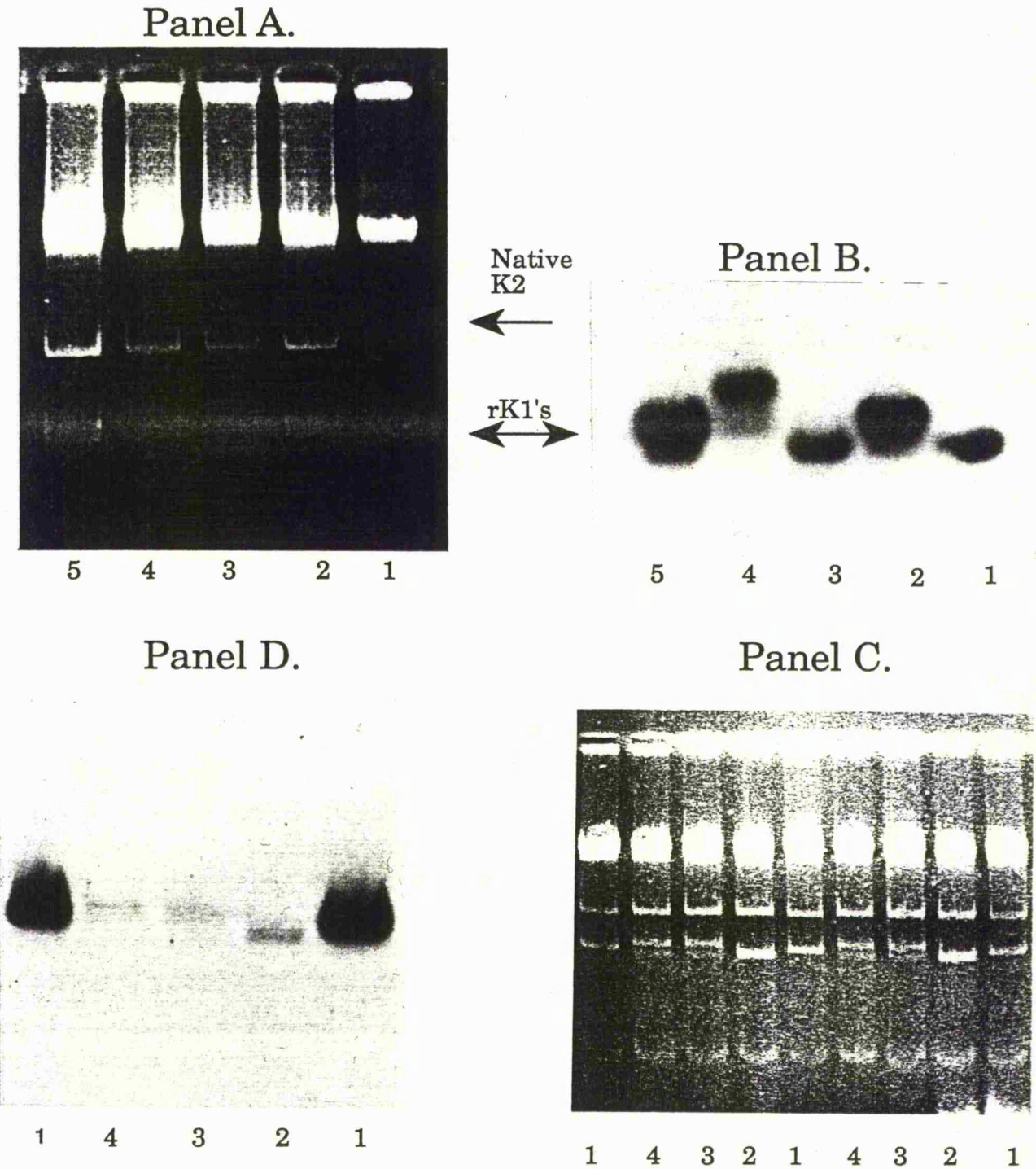
Southern analysis of recombinant K1-bearing *K. lactis* strains (see section 2.4.18) confirmed the presence of the *TRP1* gene and the size of the recombinant plasmids. Strangely after the initial transformation at least three different species of histidine tagged K1-derived plasmid, of 8.4, 8.9 and 9.4 kb, were detected (Figure 5.5A & B). This was not observed with the control transformation with the untagged gene; all the linear K1 recombinants bearing the *TRP1* gene were of uniform size (data not shown). The use of a probe corresponding to the 5' section of ORF 2 (which had been deleted in the recombinant plasmids) revealed that all trp^{+} *K. lactis* transformants retained the native K1 plasmid.

K. lactis colonies containing each of the different K1 variants were picked and were subjected to extensive sub-culturing. However, after 120 generations agarose gel analyses of the linear plasmid content revealed that all three *K. lactis* clones appeared to retain copies of the native K1. Nevertheless Southern analyses, using a 0.5 kb *Eco*R I-*Bam* H I probe corresponding to the deleted fragment of K1 ORF 2, indicated that very little native K1 was present (Figure 5.5D). Furthermore, a duplicate Southern blot using the *TRP1* probe appeared to suggest that all the K1 variants contained this marker. If so, this implies that the smaller K1 variant was generated from the larger species, due, possibly, to the deletion of the repeated ORF 1 segment. If this is the case, then one of the versions of ORF 1 is, presumably, functional. Depending upon the exact nature of the deletion, this maybe the tagged version, and if so may allow isolation of this product.

5.3 DISCUSSION.

The ability of the K-plasmids to be stably maintained and propagated in a number of yeast species, including *K. lactis*, *K. fragilis*, *S. cerevisiae* and *C. pseudotropicalis* (Gunge & Sakaguchi, 1981; Sugisaki *et al.*, 1985) may eventually allow their development as a yeast broad

Figure 5.5 Analysis of the recombinant K1-derived linear plasmids. **Panel A.** A 0.6% w/v agarose *in situ* lysis gel of five *K. lactis* JK1 TRP transformants, showing the heterogeneity of recombinant K1 plasmids. **Panel B.** Autoradiograph from the Southern blot of the gel from Panel A, probed with a radiolabelled fragment of the *TRP1* gene. **Panel C.** The 0.6% w/v agarose *in situ* lysis gel of *K. lactis* clones 1, 2 & 4 after 120 generations of growth in selective media. Lane 1 is the parental *K. lactis* JK1; lane 2 is *K. lactis* clone 1, lane 3 is clone 2 & lane 4 is clone 4. **Panel D.** Autoradiograph from the Southern blot of the *in situ* lysis gel from Panel C, of clones 1 (lane 2), 2 (lane 3) 4 (lane 4) & *K. lactis* JK1 control (lanes 1) after 120 generations in liquid media. This was probed with a radiolabelled fragment of K1 ORF 2 which had been deleted by integration of the recombinant ORF 1 and *TRP1* gene



host range vector system. The utility of these plasmids would also be facilitated by their capability for self-selection (toxin production and immunity) and product secretion. More importantly, in terms of this study, development of these plasmids as expression vectors may allow analysis of the endogenous genes, something which has, by-and-large, been difficult using 'conventional' over-expression systems.

One of the alternative techniques for isolating the putative gene products has been extensively studied in our laboratory. The reintroduction of modified versions of ORF 4 (Soond, Pers. Comm.) and ORF 6 (Schaffrath, Pers. Comm.) of K2, and ORF 1 of K1 (Ambrose, This Thesis) have all been attempted. The results have so far, however, been disappointing.

The ability to use this homologous recombination technique to modify K1 was demonstrated by the insertion of the *K. lactis TRP1* gene into ORF 2 of K1 (Soond, Pers. Comm.). The *TRP1* gene was flanked by the 3' region of ORF 1 and the 5' region of ORF2 (in plasmid pMS11) and integration of this linear fragment resulted in the reformation of the complete ORF 1 coding sequence. That the ORF 1 product was functional was demonstrated by the segregation of the native and recombinant forms of the plasmid. Taking this a step further, by modifying ORF 1, isolation of the encoded product should be possible.

A recombinant version of K1, containing a C-terminal poly-histidine tagged version of ORF 1, has been constructed. The transformation of *K. lactis* JK1 with the linear integrative cassette (Figure 5.4) containing the poly-histidine linked ORF 1 resulted in the establishment of at least two different recombinant species of K1 (Figure 5.5), in addition to native K1 (8.9 kb). One of these was slightly smaller than the native plasmid (approximately 8.5 kb) and one slightly larger (9.5 kb). Both the new 8.5 and 9.5 kb plasmids contained the *TRP1* gene, as demonstrated by Southern analyses. The largest plasmid corresponded to that expected by transformation with the cassette. The appearance of a smaller K1 derivative, however, suggested a different recombination event had taken place. This may have been due to the presence of the duplicate 3' section of ORF 1, which was present in the plasmid pMS11. This plasmid (pMS11) was not

originally designed for manipulation of ORF 1, and unfortunately, the duplicated segment was not easily removable.

Attempts were made to segregate the linear recombinant derivatives from native K1, during the course of extensive sub-culturing. After 120 generations, which should have been sufficient for complete separation of native and chimeric plasmids (Schaffrath & Soond, Pers. Comm.) very few copies of the parental K1 remained. However, two of the three clones analysed (2 and 4) appeared to retain two K1 derivatives (Figure 5.5) and both to be of a slightly different size to native K1. It is possible that the plasmids may segregate after further sub-culturing.

Although the extreme termini of polypeptides are often surface located and relatively mobile (see Chapter 6) the presence of a stretch of heterologous amino acids may interfere with the folding or conformation of the protein, resulting in the synthesis of an enzymically inactive product. Recombinant K-plasmids would, therefore, also be dependent upon the product encoded by the native gene. Such an observation was made by Schaffrath (Pers. Comm.) during manipulation of K 2 ORF 6. Modified versions of ORF 6 (bearing C-terminal epitope and chelating peptide tags) were transplaced (independently), along with the *TRP1* selectable marker, into ORF 2 of K1, and the native K1 was 'chased out'. The native ORF 6 was then disrupted with the *S. cerevisiae LEU2* gene. Although this plasmid became established in *K. lactis* strains, it could not be segregated from the native K2, despite the presence of an additional copy of ORF 6 on the recombinant K1. DNA sequencing of the PCR amplified ORF 6 failed to reveal any obvious errors. It has been proposed, in this case, that the C-terminal tags in the modified ORF 6 may interfere with the interaction of the ORF 6 and ORF 7 products (Schaffrath, Pers. Comm.).

The complete displacement of the native parental plasmid with that bearing a modified (tagged) version of a gene is, however, not sufficient to render subsequent isolation of the product straightforward. This has been demonstrated by attempts to identify and isolate both epitope and histidine-tagged K2 ORF 4 product. In a series of elegant experiments a modified version of ORF 4 from K2 was transplaced into K1 ORF 2 (Soond, Pers. Comm.).

The chromosomal *TRP1* disruption in *K. lactis* JK1 was complemented by the insertion of the *TRP1* gene, under the control of the K1 ORF 2 expression signals, into K1. Modified versions of K2 ORF 4, bearing the 9E10 or histidine 'tags' were introduced together with the *TRP1* gene. The resultant 10.1 kb recombinant plasmids were stably maintained and could be segregated from native K1. The native ORF 4 of K2 was then disrupted by insertion of the Km^R gene from Tn903, also under the control of the K1 ORF 2 expression signals. This 14.9 kb recombinant plasmid could also be segregated from native K2, but only in the presence of the recombinant K1 plasmids, demonstrating that ORF 4 of K2 was essential for K-plasmid maintenance. Extensive sub-culturing eventually allowed the isolation of prototrophic Km^R *K. lactis* colonies containing only the recombinant versions of the K-plasmids. Subsequent attempts to identify and isolate the ORF 4 product, using the 'tags' have, however, proved unsuccessful.

In retrospect, the best approach to eliminating the heterogeneity observed among the recombinant K-plasmids would probably be to remove the ORF 1 duplication. This would however, necessitate the modification of the plasmid pMS11 or one of its precursors. Furthermore, the probability of a PCR-induced error in the ORF 1 coding sequence could also be minimised by the amplification of a shorter 3' fragment of ORF 1. A 5' PCR primer containing an integral *Eco*N I site would allow insertion of the PCR sequence at a point approximately 300 bp from the end of the ORF 1 coding sequence. The subsequent direct (and complete) sequencing of this fragment prior to cloning, using (for example) the Circumvent kit, would also facilitate the avoidance of such errors. Having established a recombinant K-plasmid containing a modified version of a particular gene, and segregated this from the native version of the same gene, the identification and isolation of the required protein can still constitute a considerable problem (Soond, Pers. Comm.). This may be due to the natural very low abundance of the native product. An interesting and useful control experiment, to check the efficacy of the isolation procedure, would be to use the K-plasmids to isolate a 9E10 or histidine tagged reporter protein, possibly one of those which has already been successfully inserted into K1 or K2 (Km^R, *LEU2* or *TRP1*). The utility of this approach has also recently been demonstrated by the detection of the K2 ORF 5 product (Schaffrath, Pers.Comm.). Attempts to express this gene, with an integral epitope tag,

in *E. coli* and from a *S. cerevisiae* nuclear vector, failed. In contrast, the cytoplasmic expression of a modified ORF 5, from a recombinant K1 plasmid, was accomplished.

The native transcript of K1 ORF1 has been shown to be difficult to detect (Stark *et al.*, 1984; Romanos & Boyd, 1988). This suggests that the transcript is rare and may reflect a weak promoter. Attempts to express K-plasmid genes on recombinant K-plasmids may, therefore, be more efficient if the stronger K-plasmid promoters are utilised. The K1 ORF 2 promoter is known to be able to direct high level expression of the heterologous Km^R reporter gene, conferring resistance to 200 μg/ml kanamycin (Soond, Pers. Comm.). Northern analyses have also demonstrated that the transcript of K2 ORF 10 is the single most abundant of the K-plasmid messengers (Tommasino *et al.*, 1991). This could be due to a strong promoter, although it could also reflect intrinsic stabilising properties of the transcript. As reporter genes can now be efficiently inserted into and expressed from the K-plasmids, and suitable integrative 'cassettes' are available, there is scope for analysis of the K-plasmid promoters. As the sequences of all the K-plasmid ORF upstream conserved sequences are known, this should be relatively straightforward.

The construction of linear cytoplasmic derivatives of the K-plasmids have, therefore, highlighted a number of important points. They confirm other observations, that K-plasmid expression signals do not function in the nucleus; that K-plasmids contain regions with the ARS activity; that nuclear genes which confer a strong selective pressure cannot be expressed in the cytoplasm, under control of their own promoters. Furthermore, analyses of the *LEU2* transcript produced by *K. lactis* strains containing the cytoplasmic *LEU2* gene (Kamper *et al.*, 1991) revealed that although the transcript initiated where expected, transcription terminated at neither the nuclear *LEU2* or ORF2 terminators sequences, both of which were present in the construction, but at a point half way between the two. This demonstrates that nuclear terminators also appear not to be recognised in a cytoplasmic context.

Having developed the technology to manipulate the K-plasmids *in situ* and express heterologous genes from cytoplasmic derivatives, this system can now be utilised to analyse the K-plasmid genes themselves. This can be accomplished in a number of

different ways. The essential nature of the genes can be investigated by means of insertional disruption or by the synthesis of antisense transcripts. The latter has been attempted using a yeast nuclear vector (Chapter 7) but it would be interesting to see the effects of antisense RNAs transcribed from the K-plasmids themselves. The gene disruption approach has also been used to demonstrate that ORF 1 of K2 is not essential for K-plasmid replication, maintenance or the killer and immunity phenotype (Schaffrath *et al.*, 1992). A modified version of the *S. cerevisiae LEU2* gene, controlled by the K1 ORF 2 expression signals was inserted into K2 ORF 1 and used to confer prototrophy upon *K. lactis LEU2* auxotrophs. The recombinant K2 plasmid (15.08 kb) was larger than and could be segregated from native K2.

Interestingly Kamper *et al.*, (1991) also reported the ability, albeit at low frequency (1 transformant/μg), to isolate and retransform *K. lactis* strains with the recombinant linear plasmid pJKL1. This could, however, only be accomplished in the presence of both native K1 and K2. This suggests that recombination occurs between the exogenously added and endogenous K1, or that a pool of K1-specific terminal proteins exist in the cytoplasm. The deproteinised genome of adenovirus is known to be able, albeit at a very low efficiency, to replicate. A similar mechanism may operate with the K-plasmids. This possibility is, however, not easy to test as K1 also requires the presence of K2 for maintenance. Gel retardation analyses of cell extracts might be able to test this hypothesis. The mechanism of the observed *in vivo* homologous recombination between K-plasmids is, in itself, not clear. As recombination is usually catalysed by nuclear enzymes, K-plasmid recombination may be mediated by host encoded factors synthesised in the cytoplasm during their translocation to the nucleus, or by leakage of these proteins from the nucleus, by plasmid encoded factors, or by the K-plasmids having the ability to cross the nuclear membrane.

CHAPTER 6.

SYNTHETIC PEPTIDES AS ANTIGENS.

6.1 INTRODUCTION.

As discussed in Chapter 3 and 4, the expression of K-plasmid genes in *E. coli* has been problematic, for reasons relating to the intrinsic nature of the K-plasmid DNA. For the same reasons, the probability of over-expressing the native full length genes in *S. cerevisiae* is also remote. These two organisms remain the most versatile expression systems currently available, despite the fact that the mechanisms governing gene expression are not completely understood. Successful over-expression of full length K-plasmid genes probably depends upon the development of more 'sympathetic' systems. This may be provided by the gram positive bacterium *S. aureus*. There are however, relatively few *S. aureus* cloning and expression vectors currently available, and the construction of such would be time consuming. The emerging *in vivo* homologous recombination technology (see Chapter 5) may eventually facilitate the manipulation and over-expression of the endogenous K-plasmid genes, but is also labour intensive.

One alternative approach is to use chemically synthesised peptides to raise antibodies, which could subsequently be used to detect and isolate the K-plasmid gene products in yeast extracts. This technique has been used successfully to raise antibodies which specifically recognised the native ø29 TP (Shih *et al.*, 1983) and the adenovirus Dpol (Rekosh *et al.*, 1985).

The relatively recent interest in the antigenic properties of synthetic peptides (SP), short chemically synthesised linear chains of amino acids, results from their ability to mimic antigenic sites in whole proteins (Lerner, 1982; Sutcliffe *et al.*, 1983). This was first demonstrated using the enzyme lysozyme (Arnon *et al.*, 1971). The availability of small peptide fragments with the ability to elicit an immune response, producing antibodies which specifically cross-react with the complete protein, could significantly simplify the isolation and characterisation of previously unavailable proteins, such as those putatively encoded by the K-plasmids. The use of conventional biochemical purification techniques for such proteins may be problematical due to a lack of suitable assays or extraordinarily low natural abundances. Furthermore, the use of recombinant DNA techniques does not necessarily guarantee successful over-production of a protein (see Chapters 3 and 4).

The development of solid phase peptide synthesis technology has made the synthesis of short peptides an increasingly attractive alternative to the generation of small protein fragments by either proteolytic or chemical cleavage, both of which are complicated by problems with contaminants. SPs have already found a major application in the development of vaccines eg. Foot and Mouth Disease (Bittle *et al.*, 1982; DiMarchi *et al.*, 1986) and Malaria (Patarroyo *et al.*, 1987; Tam *et al.*, 1990). Consequently, much of the work undertaken to date concerning the selection and antigenicity of SPs has been in relation to their ability to protect immunised animals from viral infection. Anti-peptide antibodies have also been used to confirm the occurence of proteins predicted to exist *in vivo* by DNA sequence data (Sutcliffe *et al.*, 1980). The utility of SPs has recently been examined, in considerable detail, by Koprowski & Melchers (1986), Van Reggenmortal (1988) and Grant (1992).

The antigenic reactivity of a protein refers to its capacity to react specifically with the binding sites of certain immunoglobulins. The portion of an antigen that comes into contact with these antibody binding sites, is the epitope. The complex nature of protein antigens means that each may have multiple epitopes. In fact, the entire accessible surface of a protein can be regarded as consisting of a large number of overlapping epitopes. Two classes of epitope can, generally, be distinguished on the basis of their structural composition, at the primary sequence level. Continuous (or sequential) epitopes are defined as those which are composed of a stretch of residues which are continuous in the primary sequence. In contrast, discontinuous (or conformational) epitopes are composed of residues which are not contiguous in the primary sequence, but are juxtaposed as a result of secondary, tertiary or quaternary structural features (ie. folding of the peptide chain or positioning of separate peptides). In reality, this distinction is not so clear cut, as continuous epitopes can also often be regarded as small parts of larger discontinuous epitopes. An epitope is, therefore, a structural feature of the antigen, and as such the antigenic reactivity of a protein is not the same in its native and denatured state. The recognition of a protein by antibodies raised to SPs is dependent upon the state in which the protein is presented. This can, for example, be influenced by the binding of protein to plastic surfaces during immunoassays or antigen presentation by lymphocytes during the

initial immune response. Antibodies raised to the native protein or to small peptides corresponding to discontinuous epitopes may, therefore, not recognise the denatured form of the protein. The disruption of the four disulphide bridges which maintain the conformation of hen egg white lysozyme has been observed to result in complete loss of antigenic reactivity with antibodies raised to the native protein (Arnon, 1986; cited in Arnon, 1989). Antipeptide antibodies do, however, generally tend to cross react strongly with the denatured form of the protein. This is probably because unfolded proteins exhibit a greater structural resemblance to the epitopes presented by the majority of SPs. The discontinuity of most epitopes may also be reflected by the fact that antibodies raised to SPs often only weakly cross-react with the native protein, highlighting the limited structural resemblance. Such antibodies may, however, also recognise discontinuous epitopes in the native protein, as a result of the peptide mimicking part of the complete epitope structure.

The major problem encountered when raising antibodies to SPs is whether the antibodies will recognise the complete protein. Currently, the only structural information available with the majority of proteins is the primary sequence, which is often simply derived from DNA nt sequence data. The successful application of SP technology is, therefore, limited by the ability to predict the location of antigenic sites in a protein's three dimensional structure directly from its amino acid sequence. There have been concerted attempts to formulate reliable 'models' which enable a polypeptides sequence to be related to its antigenicity. Detailed analyses of the antigenic properties of well characterised proteins (primarily lysozyme, influenza haemagglutanin, tobacco mosaic virus protein and myoglobin) has facilitated the development of a variety of different computer models (algorithms), the data from which has been extrapolated to the primary sequences of other proteins. A considerable number of structural parameters have subsequently been shown to correlate with the location of known continuous epitopes and many conflicting claims have appeared in the literature as to which are the most reliable methods. The major features of protein structure which are known to correlate with antigenicity, and which, if targeted for peptide synthesis, may increase the probability of antigenicity include hydrophilicity, amphipathicity, segmental mobility, sequence variability and static surface accessibility. As yet, none of these have been tested sufficiently to determine their

predictive significance, but the results to date have revealed that no single method is consistently successful (Van Regenmortel, 1989).

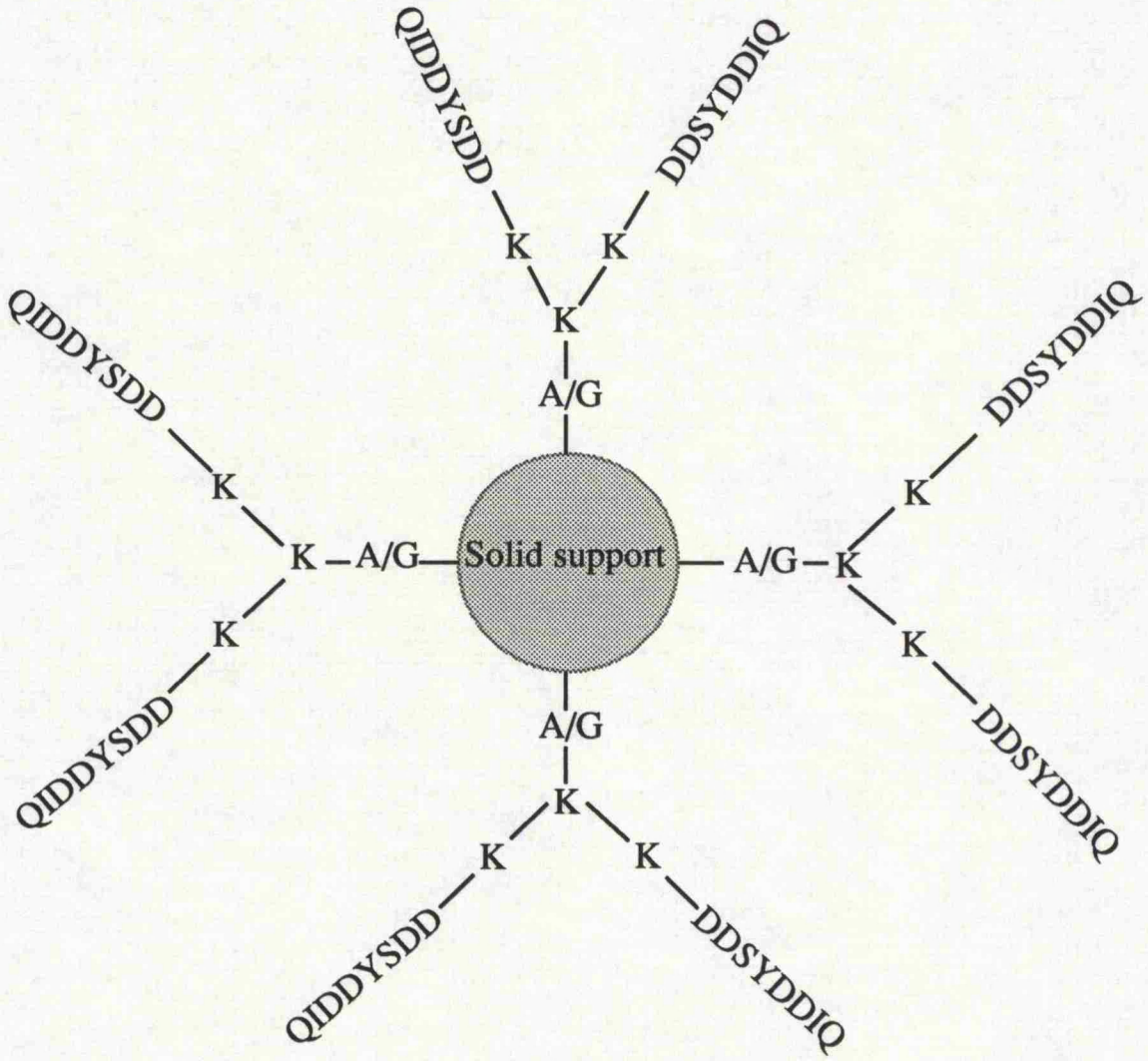
The smallest peptides that consistently elicit the generation of antibodies which recognise the complete protein are of 6 residues. Generally peptides of 6-15 residues are synthesised, but the immunogenic response becomes progressively weaker as the length of the peptide decreases. Due to their small size (an average ten residue SP having a molecular weight of approximately 1 kDa) these peptides may not be immunogenic on their own. In order to induce an immune response, a peptide must contain binding sites for both B- and T-lymphocytes. Most peptides do not contain these, so their immunogenicity is boosted by chemical coupling to carrier proteins, prior to injection. This has been one of the major drawbacks with the use of SPs in vaccine development. Carrier molecules are often highly antigenic and the relatively large mass of the carriers, compared to the peptides, may result in a low ratio of anti-peptide to anti-carrier antibodies. Thus only a small proportion of the total antibodies synthesised are of the required specificity. This led to the development of multiple antigen peptide (MAP) resins (Tam, 1988). Here a small non-immunogenic multivalent lysine polymer core forms a branching scaffold onto which peptides can be directly synthesised (Figure 6.1). These cores can, for example, be constructed with 4, 8 or 16 branches. The advantage of these macromolecules is that they dispense with the need for carrier proteins.

The emerging SP technology has great potential in the development of vaccines, and at a more fundamental level, to enable the detection and isolation of proteins. This approach might, therefore, present a means of detecting and isolating K-plasmid gene products. With this is mind we decided to use this technique, in addition to two relatively recent developments in the generation of antibodies; the use of MAPs to replace carrier proteins and to raise the antibodies in chickens.

6.2 RESULTS.

The use of SPs is currently limited by the inability of researchers to predict, with any certainty, antigenic sites in proteins from amino acid sequence data. This is the only information available in a great many cases. As none of the methods currently available to

Figure 6.1 The general structure of the multiple antigen peptide (MAP) SP1, containing the most N-terminal sequence of the two selected potential K2 ORF 2 epitopes (QIDDYSDD). The MAP SP2 differed only in the sequence of the terminal SP (TYGDINKD). A/G represents alanine or glycine residue spacers, introduced to minimise steric hinderence during peptide synthesis. The single letter code for amino acids is used.



identify immunogenic sites in a protein's primary sequence are particularly reliable, the best practical approach for maximising the probability of selecting antigenic sites, is to use a combination of these models. The target sequence can then be defined as that which yields maximal values with the majority of models.

6.2.1 Analyses Of K1 And K2 DNA Polymerases For Potential Epitopes.

The deduced amino acid sequences of the putative K1 and K2 Dpols were analysed for the location of potential epitopes by Dr. M. Crossman (The Advanced Biotechnology Centre, Charring Cross and Westminster Medical School, London), on the basis of several criteria: Hydrophilicity (Kyte & Doolittle, 1982), Antigenicity Index (Jameson & Wolf, 1988), Surface Probability (Emini, 1985), Secondary Structure (Chou & Fasman, 1978; Garnier *et al.*, 1978), and Chain Flexibility (Karplus & Schulz, 1985). In view of the fact that much of the previous work, on the heterologous expression of K-plasmid Dpols, had focused upon the putative K2 encoded protein, this enzyme was selected as the target for peptide synthesis. Upon examination of the resultant data, and on the advice of Dr. Crossman, two specific sites were selected in the N-terminal portion of ORF 2 of K2, where the maximal values for each of these criteria coincided. Two SPs, each of eight residues, corresponding to residues 11-18 (SP1, QIDDYSDD) and 184-191 (SP2, TYGDINKD) were therefore synthesised. Both sequences fall within regions predicted to have high relative flexibility, surface probability, hydrophilicity, antigenicity index and secondary structure. Some of the data regarding these two sites are shown in Table 6.1.

The two SPs corresponding to the selected target sites were synthesised as tetrameric MAPs. This was the first time that this technology had been used at Leicester University. The resulting complexes had molecular weights of approximately 5 kDa, the central lysine core constituting about 25% by mass of the final product. The final proportion, of course, depends on the 'valency' of the core (ie. the number of branches) and the length of the peptide synthesised; the greater the number of branches and the longer the peptide, the smaller the final proportion represented by the core. Increasing both branch number and peptide length may also increase the antigenicity of the complex. However, there have been problems with product solubility associated with the 8-branch resins supplied by

Table 6.1. A selection of the K2 ORF 2 analyses data provided by Dr. M. Crossman. The choice of the synthetic peptide sequences was based upon regional maxima in this data. The positions marked * below show the two sequence selected for synthesis of peptides. Hydrophilicity was predicted according to the method of Kyte-Dolittle (1982); surface probability according to Emini (1985); flexibility according to Karplus & Schulz (1985); secondary structure according to Chou & Fasman (1978); antigenicity index according to Jameson & Wolf (1988). In terms of secondary structure 'H' means predicted location in α-helix, 'T' in turn and 'B' in β-sheet.

Pos	Residue	Hydrophilicity.	Surface probability	Flexibility	2ndary structure	Antigenicity index
7	F	-0.629	0.294	0.96	H	-0.6
8	L	-0.629	0.617	0.982	H	-0.6
9	N	-0.729	0.428	1.012	-	-0.45
10	S	0.029	0.826	1.044		0.45
11*	**Q**	**0.929**	**1.673**	**1.054**	**-**	**1.3**
12*	**I**	**1.657**	**1.63**	**1.053**	**-**	**1.3**
13*	**D**	**1.271**	**1.63**	**1.052**	**T**	**1.5**
14*	**D**	**1.657**	**1.571**	**1.042**	**T**	**1.7**
15*	**Y**	**1.657**	**3.744**	**1.037**	**T**	**1.7**
16*	**S**	**1.657**	**1.571**	**1.039**	**T**	**1.3**
17*	**D**	**1.657**	**1.63**	**1.031**	**T**	**1.3**
18*	**D**	**0.757**	**0.901**	**1.019**	**T**	**0.95**
19	I	0.029	0.554	0.998	H	0.3
20	E	0.557	0.65	0.977	H	0.6
180	P	0.043	0.432	0.964	B	0.3
181	I	-0.5	0.312	0.957	B	-0.6
182	I	-0.871	0.658	0.964	B	-0.6
183	D	-0.214	0.421	0.992		-0.3
184*	**T**	**0.057**	**1.003**	**1.013**	**T**	**1**
185*	**Y**	**0.057**	**1.003**	**1.023**	**T**	**1**
186*	**G**	**1.2**	**0.966**	**1.032**	**T**	**1.35**
187*	**D**	**1.257**	**1.338**	**1.035**		**1.3**
188*	**I**	**1.657**	**1.426**	**1.038**	**T**	**1.1**
189*	**N**	**1.071**	**1.248**	**1.039**	**T**	**1.5**
190*	**K**	**1.2**	**1.171**	**1.032**	**T**	**1.7**
191*	**D**	**1.2**	**2.686**	**1.016**	**T**	**1.7**
192	F	2.4	3.341	0.999	H	0.75
193	Y	1.257	1.171	0.994	H	0.75
194	N	1.257	1.402	1.004	H	0.9

Applied Biosystems Inc. (Lilley & Seddon, Pers. Comm.). Peptide synthesis is, importantly, also relatively expensive. During the synthesis of the two peptides, further problems emerged, mainly concerning the poor sequential coupling efficiencies at each stage, as reflected in the low final weight gain. Analysis of the final products by mass spectroscopy, however, confirmed the correct structure (Seddon, Pers. Comm.).

6.2.2 Chickens As Hosts For Polyclonal Antibody Generation.

The generation of antibodies has, generally, been accomplished by the immunisation of rodents. The resulting antisera is subsequently collected, 2-4 months later by 'bleeding' the animal. There has, however, been an increasing awareness of the advantages of using chickens to raise polyclonal antibodies. As the full spectrum of the hens serum antibodies are secreted into the yolk of the eggs, these provide an excellent source of antibodies. Furthermore, this simplifies harvesting and subsequent purification, and importantly, is non-invasive. The differential PEG precipitation allows the simple purification of antibodies to greater than 90% (Gassman *et al.*, 1990). Additionally, much smaller amounts of antigen are required to elicit an immune response (two injections of 10-20 μg are generally sufficient). Specific antibodies typically appear within 20 days of the first immunisation, attaining a maxima after 25-30 days, the level remaining high for several months. The use of Freunds Complete Adjuvant, which is responsible for the very high titre of yolk antibodies and the prolonged duration of the immune response, also does not produce an inflammatory response (as it does in rodents).

Polyclonal antibodies raised in chicken egg yolk, in response to immunisation with the two SPs, should be ideal for detection of the native K-plasmid encoded Dpols by Western analyses. This combines the high resolution of gel electrophoresis with the high specificity and sensitivity of immunochemical detection. Apart from detecting the presence of the protein, this may also provide information concerning its size and abundance. The peptides were therefore used to immunise chickens and yolks from sets of four eggs were harvested 25, 30, 35 and 40 days after the initial immunisation.

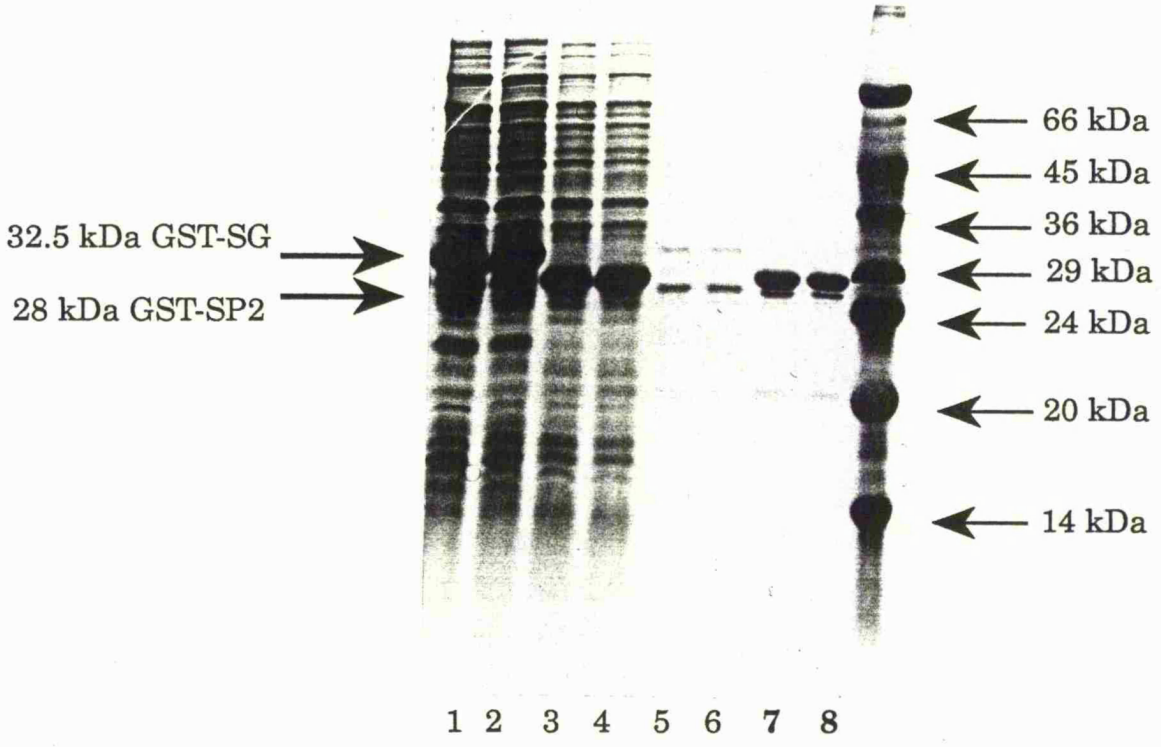
6.2.3 Western Analyses Using The Chicken Antibody Samples.

Positive control samples for the antisera raised to SP1 and SP2 were provided by the nitrocellulose immobilised MAP. A further control for SP1 was provided by the use of the 32 kDa K2 Dpol-GST fusion protein produced from the plasmid pCA30 (see Chapter 4). A similar control protein for SP2 was also constructed. Two oligonucleotides, CA19 and CA20 (Appendix 2B) corresponding to both strands of the SP2 DNA sequence were synthesised, incorporating terminal *Eco*R I and *Bam*H I half sites. These were annealed and cloned directly into the *Bam*H I-*Eco*R I sites of the plasmid pGEX2T (yielding plasmid pCA31). Cell extracts prepared from *E. coli* strains containing this plasmid were analysed by SDS-PAGE and revealed that, upon IPTG induction, the GST-SP2 protein was produced to at least 10% of total cellular protein. This fusion protein was fully soluble, and could be isolated by affinity chromatography (Figure 6.3). As expected, since the fusion contains only eight additional residues, the isolated recombinant protein was indistinguishable, on the basis of its electrophoretic mobility, from the non-fusion GST molecule.

Samples of these antibody preparations were then used in immunoblot (Western) analysis of the positive control GST fusion proteins, to confirm that the peptides had been immunogenic, prior to attempts to detect the native K2 Dpol in crude yeast extracts. The yolk antibody samples were used as the primary antibody, at dilutions ranging from 0.1 to $5x10^{-4}$ v/v, in conjunction with goat anti-chicken secondary and rabbit anti-goat-peroxidase tertiary antibodies.

Crude cell extracts from the GST-Dpol overproducing strains containing plasmids pCA30 and pCA31 and affinity purified protein expressed from pCA31, were electrophoresed on denaturing 15% w/v polyacrylamide gels and electroblotted (Figure 6.3). The transfer of protein to nitrocellulose was found to be relatively efficient, as judged by Ponceau S staining of the membrane after blotting. However, at no stage was any reaction to these proteins observed by Western analysis. Furthermore, the immobilisation of samples (10 µl, containing from 2 ng to 2 µg) of the MAPs directly to the membrane and immuno-

Figure 6.3 15% SDS-PAGE of cell extracts from IPTG induced *E. coli* NM522 bearing the over-expressing plasmids pCA30 and pCA31. pCA30 directs the synthesis of GST with a C-terminal fusion of the synthetic gene fragment (see Chapter 4) and contains the sequence corresponding to SP 1 (QIDDYSDD). pCA31 expresses GST with a C-terminal fusion of the sequence corresponding to SP 2 (TYGDINKD). Duplicates of this gel were electroblotted (for use as positive control reactions for antibody production) and were probed with antibody samples, prepared from hen egg yolks, in response to immunisation with SP 1 & 2. Lanes 1-4, total cellular protein; lanes 5-8, affinity purified protein. Samples were prepared from strains carrying pCA30 (1, 2, 5 & 6) and pCA31 (3, 4, 7 & 8).



analysis also failed to reveal any antibody binding to the membranes. The MAPs, therefore, appear not to have been immunogenic.

6.3 DISCUSSION.

The antigenicity of proteins is largely determined by the three dimensional arrangement of their polypeptide(s). In the absence of detailed data concerning the tertiary structure of a protein, however, the search for potentially antigenic sites in a protein is restricted to the identification of surface located continuous epitopes. The use of SPs in the generation of antibodies which specifically recognise the complete native protein is currently limited by the inability to accurately predict the position of such epitopes. This stems from the fact that the only structural information available concerning most proteins is the primary sequence. A variety of methods to identify features in the primary structure which correlate with exceptional immunogenicity have been formulated (see below), but none have proven to be consistently reliable.

Analysis of protein structure has shown that hydrophobic amino acids tend to be buried within the tightly packed native protein core, whereas hydrophilic residues are often located on the exterior where they interact with water. The ability to evaluate hydrophobic/hydrophilic tendencies of an amino acid sequence may, therefore, facilitate the prediction of whether peptide regions are buried or exposed (Kyte & Doolittle, 1982; Hopp & Woods, 1983) and, therefore, antigenic. The method of Hopp and Woods (1985) was used, successfully, to predict the locations used for peptide synthesis with the human *c-myc* protein (Evan *et al.*, 1985).

Epitope probability predictions, based on secondary structure models for α-helix (Chou & Fasman, 1978) and β-sheet (Garnier *et al.*, 1978) formation, tend to exhibit inverse correlations with maxima in hydrophilicity, due to the fact that both of these structures tend to occur in the protein core. On the other hand, the presence of proline residues has also been proposed to be of predictive value, due to their involvement in the establishment of surface located β-turns, which often form part of known epitopes. Furthermore, the development of computer algorithms to measure the degree of amphipathicity within a secondary structure (which arises from the periodicity of hydrophobic residues within the primary

sequence, and results in one side being appreciably more polar than the other) have predicted that antigenic sites tend to be located within such regions.

As antibodies can only bind to the surface structures of native proteins, certain protein segments may be particularly antigenic due to their exceptional surface exposure. NMR and X-ray crystallographic analyses have also demonstrated that regions corresponding to known epitopes are often located in relatively mobile portions of a protein, and these often coincide with highly accessible segments at the proteins surface (Tainer *et al.*, 1984). Thus the prediction of the more mobile portions of a protein, increases the probability of these being antigenic (Westof *et al.*, 1984; Karplus & Schultz, 1985). In the majority of cases studied, a proteins N- and C-termini are located at the molecules' surface and in close proximity to each other. These peptide termini are also less constrained than internal segments and therefore have high relative flexibility. Such features are probably responsible for the fact that these regions often correspond to continuous epitopes. This may explain why antibodies raised to the whole protein often cross react better with short terminal peptides than internal peptides. SPs corresponding to the terminal 10-15 residues are, therefore, likely to be suitable immunogens and many groups routinely used this approach, with good results. Antibodies which recognised the native ø29 TP (Shih *et al.*, 1983), adenovirus Dpol (Rekosh *et al.*, 1985) and the unknown gene product of Moloney Leukaemia virus (Sutcliffe *et al.*, 1980) were raised using this method.

Regions of high sequence variability, in families of homologous proteins, also often correspond to the location of epitopes. Such regions, where local conformational changes are most tolerated, are probably surface located and are not likely to be involved in long range interactions that stabilise the secondary or tertiary structure. These sequences have been successfully targeted and often correspond to structural features in viral proteins (eg. Foot and Mouth Disease virus) which are involved in evading the host's immune system (Bittle *et al.*, 1982).

These requirements, however, may not be necessary in the design of SPs for use in identification or isolation of unfolded or denatured proteins. In theory, SPs corresponding to any amino acid sequence can be used to raise antibodies. These should then be able to

recognise the completely unfolded form of the protein, although they may or may not recognise the native protein, depending upon the location of the selected sequence in the three dimensional structure. The importance of targeting surface exposed regions lies in the detection and isolation of native proteins. There is no reason, *per se*, why the two sites selected in this case should be any more or less antigenic than any other sequence in the protein. The inability of these SPs to elicit an immune response, as demonstrated by the positive control reactions, lies not with the sequences themselves but with associated factors. Such reasons include the structure being too small (each peptide being too short or the core not having enough branches), poor coupling of the peptides to the cores, or these MAPs simply not being immunogenic in chickens. In all of the cases reported to date, rodents have been used to raise antisera against MAPs.

The use of MAPs has an enormous potential to eliminate the need for hazardous adjuvants, which are applied to increase the efficiency (rapidity and duration) of the immune response, by the incorporation of additional epitopes to stimulate T and B-cells (Tam *et al.*, 1990). Furthermore, the synthesis of multiple peptides corresponding to different organisms could facilitate the development of vaccines which allow the simultaneous immunisation against several different diseases. The central MAP lysine polymer is small, less than 900 Da, and the synthesis of multiple peptides on a single core results in the required epitope constituting a very high proportion of the final complex. Six peptides of 9-16 residues, synthesised on octomeric cores, were all reported to be strongly immunogenic, and five recognised the cognate sequences in the native protein (Tam, 1988). Similarly, Wang *et al.* (1991) used an octomeric 33 residue peptide corresponding to a highly variable region of the HIV-1 coat protein, to raise antisera which neutralised the virus in guinea pigs. The synthesis of a 14 residue peptide, corresponding to the β-chain constant region of the human T-lymphocyte antigen receptor, on an octameric scaffold (Posnett *et al.*, 1988) resulted in the core only constituting 7% (by mass) of the 14.3 kDa product. The product was highly immunogenic, and the resulting polyclonal and monoclonal antibodies not only recognised the linear monomeric form and the polycentric form, but also the native protein.

Interestingly, the outer terminal residues of the octomeric peptides were the most antigenic (Posnett *et al.*, 1988). The removal of these residues resulted in progressive weakening of the affinity of antibodies raised to the full length SP. This reinforces the importance of the appropriate orientation of peptide linkage to the carrier. Some monoclonal antibodies raised to the octomeric peptides did not, however, recognise a linear monomeric peptide of the same sequence. This presumably occurred due to the high uniform density of the peptides which effectively resembled discontinuous epitopes. This also supports the suggestion that continuous epitopes may simply form part of larger discontinuous epitopes.

In this case, the length of the peptides, coupled with the number of peptides per core may have been insufficient to elicit an efficient response. Despite protestations to the contrary by the manufacturer (Applied Biosystems) and the reports cited above there have, to date, been relatively few reports of the successful application of the poly-lysine resins. Similar results, highlighting the current difficulties being experienced with MAPs, have been reported by other researchers (Dr. I. Moss, Pers. Comm., The Advanced Biotechnology Centre, Charing Cross and Westminster Hospital, London). Furthermore, there have been, to my knowledge, no reports of the successful use of the tetravalent MAP resins. The possibility exists that MAPs synthesised on tetravalent cores may not be as antigenic as would be if synthesised on octavalent cores. The two peptides synthesised in this case were small, only 8 residues long, the complete molecule having a molecular weight of only 5 kDa (approximately). The synthesis of longer, 10-15 residue, peptides may have aided the development of an immune response. To date all reports of the production of antigenic octameric cores reveal that the peptide length varied from 9 to 33 residues on each branch. Furthermore, during the synthesis of these MAPs, a low weight gain was observed, suggesting a low coupling efficiency (Seddon & Lilley, Pers. Comm.). This may reflect the possibility that many of the lysine branches were still empty.

The available data therefore suggest that the two SPs, constructed in the form of MAPs, did not induce an immune response upon injection into the chickens. In retrospect, a better approach may have been to synthesise peptides, of at least 10-15 residues, corresponding to both chain termini, as well as the two internal sites selected. The use of a single 18

residue N-terminal peptide would also have incorporated the first internal sequence which was selected (SP1). A useful additional control experiment would have been to synthesise these peptides in the conventional monomeric linear form and then couple these to a carrier protein. This would, at least, have confirmed the suitability of chickens for raising antibodies against MAPs.

CHAPTER 7.

ANTISENSE REGULATION OF K-PLASMID GENE EXPRESSION.

7.1 INTRODUCTION.

7.1.1 Natural Antisense Systems.

7.1.2 Manipulation Of Antisense Systems For Gene Analyses.

7.2 RESULTS.

7.2.1 Construction Of A *K. lactis* -*E. coli* Shuttle Antisense Vector.

7.2.2 Cloning Of A K-plasmid Antisense RNA Gene.

7.2.3 Expression Of The K-plasmid Specific Antisense Gene.

7.3 DISCUSSION.

7.1 INTRODUCTION.

Recombinant DNA technology has enabled the isolation of almost any gene, and the construction of gene libraries for almost any organism. To ascertain the function of a gene is, however, more difficult. Antisense (AS) technology may, therefore, have applications in the characterisation of gene product function. This could be achieved by their specific inhibition and analysis of the resultant phenotypes. For example, inhibiting the expression of the HeLa cell initiation factor eIF-4E gene, the product of which is involved in the mRNA binding to the ribosome, using a small AS RNA was lethal (Benedetti *et al.*, 1992). Furthermore, AS molecules have an enormous medical and agricultural potential in the specific inhibition of infectious viral, bacterial and fungal diseases, some of which have been recently discussed by Weintraub (1990).

The K-plasmids contain at least 14 genes, transcripts to each of which have been identified (Stark *et al.*, 1984; Romanos & Boyd, 1988; Fleming & Meacock, unpublished results). Of these, only three genes (K1 ORFs 2, 3 and 4) have been assigned definite functions in the toxin and immunity phenotype. The functions of the remaining 11 ORF products are only tentative, and are based on amino acid sequence similarities to previously identified proteins. For example, the products of K1 ORF 1 and K2 ORF 2 are thought to be Class B Dpols. The function of these products could be elucidated, indirectly, by specific disruption of the individual genes and analyses of the effects upon plasmid replication and maintenance. Until recently, however, this has proved difficult due to the inability to manipulate these plasmids. The selective use of AS RNA technology may, therefore, present a viable alternative to direct manipulation, as a tool for analysing the function of K-plasmid genes.

7.1.1 Natural Antisense Systems.

Gene expression in both prokaryotes and eukaryotes is controlled by the products of regulatory genes, in the form of repressor and activator proteins. Regulatory genes have now been discovered that do not encode protein products, but untranslated RNAs. Such RNAs are complementary to, and interact directly with, their mRNA or DNA targets. The result of the specific nt base pairing with the corresponding coding or 'sense' sequence, is inhibition of gene expression. The existence of natural AS regulatory systems was first

discovered in the *E. coli* ColE1 and FII-type plasmids (reviewed in detail by Euguci *et al.*, 1991; Cesareni *et al.*, 1991). Replication of these DNA episomes is negatively controlled by untranslated RNAs, which not only regulate plasmid copy number, but also inhibit the stable maintenance of two similar plasmids in the same cell, a phenomenon known as plasmid incompatibility (Tomizawa & Itoh, 1981). Antisense RNAs have now also been implicated in Insertion Sequence transposition, regulation of prokaryotic gene expression and phage reproduction. These, together with some of the potential applications of AS technology, have been reviewed by Weintraub *et al.*, 1985; Green *et al.*, 1986; Simons, 1988; Toulme & Helene, 1988; Weintraub, 1990; Euguchi *et al.*, 1991. However, the existence of naturally occurring AS regulatory systems has not, as yet, been conclusively demonstrated in eukaryotes, although the presence of potential AS promoters has been reported, eg. the intron of the *S. cerevisiae* actin gene (Thompson-Jager & Domdey, 1990).

The first naturally occurring cellular AS system was identified during the characterisation of the *E. coli* outer membrane porins, OmpC and OmpF, and their mode of osmoregulation (Mizumo *et al.*, 1984; Aiba *et al.*, 1987; Andersen *et al.*, 1989). The expression of the *E. coli* cyclic AMP receptor protein (*crp*) gene is now also known to be autoregulated, by an AS RNA which is synthesised in the presence of cAMP (Cossart & Sicquel-Sanzey, 1985; Okamoto & Freundlich, 1986). The transposon IS10 encodes the protein responsible for its own transposition (Halling *et al.*, 1982). The gene for this transposase (*tnp*) lies within one of the ITRs, but expression of this gene is negatively regulated by a small AS RNA transcribed from within the same ITR.

7.1.2 Manipulation Of Antisense Systems For Gene Analyses.

The application of AS technology to the disruption of gene expression is becoming increasingly broad. Viral systems have been particularly amenable to such inhibition, and of these, retroviruses are amongst the most promising targets. The presence of high steady state levels of AS transcripts *in vivo* may allow the development of 'intracellular immunity'. The conversion of the ss RNA genome into a ds cDNA copy during the retroviral replication cycle may be particularly susceptible. For example, the synthesis of the Moloney Murine Leukaemia and Avian Myeloblastosis Virus genomic cDNA by the

virally encoded reverse transcriptase was inhibited by a small AS RNA which bound to the 3' end of the genome, rendering the RNA hybrid susceptible to the reverse transcriptase-associated RNase H (Boiziau *et al.*, 1992). This cleavage effectively blocked synthesis of the viral DNA genome and hence its subsequent integration into the host chromosome. Of course, HIV is currently one of the most intensely studied retroviral systems and AS technology may have great potential here (Rhodes & James, 1990; Rhodes & James, 1991; Rittner & Sczakiel, 1991; Sczakiel & Pawlita, 1991).

Antisense inhibition has also been used to study *E. coli* gene expression. The insertion of a 5' fragment of the *E. coli lpp* gene, in the reverse orientation into an *E. coli* expression vector, led to high levels of an AS RNA being synthesised (Coleman *et al.*, 1984). This was reflected by a sixteen-fold reduction in *lpp* synthesis within five minutes. Doubling the AS gene dosage was subsequently found to double the efficiency of inhibition. As the functional half life of the *lpp* transcript was known to be approximately twelve minutes, this suggested that the AS RNA inhibited the translation of the existing mRNA.

The production of AS RNA in eukaryotic systems has been approached in three main ways: by the direct micro-injection of AS RNA (eg. with β-globin in *Xenopus* oocytes); the micro-injection of transient expression plasmids (eg. the HSV thymidine kinase gene in tissue culture cells); or by the stable transformation with AS genes (eg. with plants or tissue cultures). In fact, the first example of the use of an artificial AS system to specifically modulate gene expression involved the micro-injection of an AS HSV thymidine kinase gene into mouse cell cultures (Izant & Weintraub, 1984; Kim & Wold, 1985). The micro-injection of AS RNAs corresponding to the extreme 5' end of the β-globin mRNA, into *Xenopus* oocytes also specifically inhibited β-globin expression (Melton, 1985).

There has been a rapidly increasing number of AS applications in plants, many of which have great commercial potential. The resistance of crop species to viral infection is one such target. For example, the replication of Tomato Golden Mosaic virus has been blocked in transgenic tobacco plants (Day *et al.*, 1991). Endogenous plant enzymes and hormones have also been targeted. Production of the gaseous plant hormone ethylene, which is responsible for fruit ripening and abscission, in tomato plants was specifically

reduced in a dose-dependent manner, by upto 67% using of a 1.1 kb AS RNA gene (Hamilton *et al.*, 1990). The enzymes polygalacturonidase and β-galacturonidase (GUS) are involved in the softening of the plant cell wall and therefore contribute to the spoiling of fruit and vegetables; inhibition of these enzyme could significantly increase the shelf life of such produce. Endogenous enzyme activities have been reduced by up to 90% in both tomato (Smith *et al.*, 1988) and tobacco (Robert *et al.*, 1989) by AS RNAs.

Antisense technology may, therefore, offer an ideal way to specifically reduce and in some cases completely inhibit, gene expression. This technique also provides a powerful tool for analysing the function of gene products. As such, AS modulation of gene activity may yield some insight in to the function of K-plasmid genes.

7.2 RESULTS.

7.2.1 Construction of a *K. lactis* Antisense RNA Vector.

Mitotically stable and regulated expression vectors, which direct high level gene expression, are available for use in *S. cerevisiae*. Many of the features necessary for these properties are also functional in *K. lactis*. Some, however, are not, but the corresponding *K. lactis* elements have now been cloned. The construction of a suitable vector from which to express a K-plasmid specific AS RNA in *K. lactis* was accomplished by a fusion of the plasmids pG1 (Schena *et al.*, 1991) and KEp6 (Wesolowski-Louvel *et al.*, 1988).

Autonomous replication in *K. lactis* is, of course, essential. The 2μ *ori* is not functional in *K. lactis*, and although the K-plasmids themselves do contain sequences which confer these properties upon nuclear plasmids, these were not exploited. Instead this was accomplished by the inclusion of the pKD1 *ori*. pKD1 is a small nuclear circular plasmid which was originally isolated from *K. drosophilarum*, but has been transferred to, and is stably maintained in *K. lactis* (Chen *et al.*, 1986). In many respects, pKD1 is equivalent to the 2μ plasmid of *S. cerevisiae*. The strong constitutive *S. cerevisiae GPD* promoter and *PGK* transcriptional, translational terminator and polyadenylation signals, all of which are functional in *K. lactis*, were used (Schena *et al.*, 1991). The *GPD* promoter did not contain an initiating codon, but as translation of the RNA was not required, this was not important. Selectable markers were provided by the *S. cerevisiae TRP1* and *URA3* genes, both of

which are also functional in *K. lactis*. However, *K. lactis URAA* mutants exhibit a relatively high reversion frequency, so this gene was not used for selection purposes.

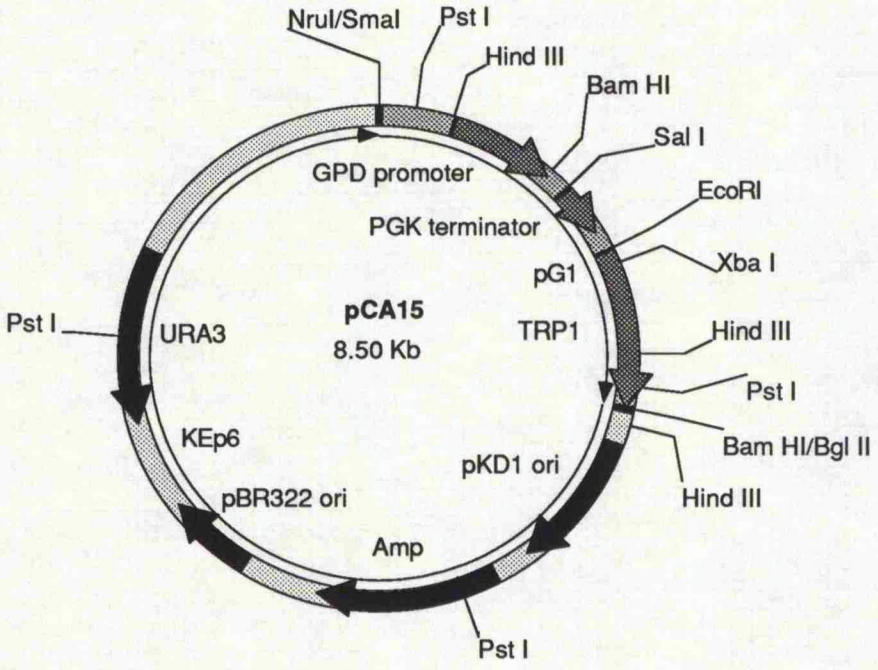
The construction itself was relatively simple. A 2.5 kb *Bgl* II-*Sma* I fragment of plasmid pG1, containing the expression signals and the *TRP1* marker, was cloned between the *Nru* I-*Bam* H I sites of plasmid KEp6, yielding plasmid pCA15. This resulted in the deletion of the central portion of the tetracycline resistance gene in KEp6. The structure of the AS vector, pCA15 is shown in Figure 7.1.

7.2.2 Construction of the Antisense Gene.

The plasmid K1 is known to be dependent upon K2 for replication and maintenance. Despite this, the functions of the majority of the K2 genes remain unknown. The inhibition of any K2 gene, except ORF 1 (Schaffrath *et al.*, 1992) may, therefore, result in the non-specific loss of both plasmids. Both ORF 2 of K2 and ORF 1 of K1 are thought to encode Dpols. Presumably, these enzymes are plasmid specific, or one gene is redundant. Analyses of mutant K1 plasmids, the lack of any K2 mutant plasmids and the detection of both the putative Dpol transcripts (Romanos & Boyd, 1988; Fleming & Meacock, unpublished results), however, suggests that both enzymes are essential. The K1 Dpol ORF is thought to be the only gene on K1 involved in plasmid replication, the others encoding toxin and immunity determinants. Antisense inhibition of the K1 DNA polymerase gene would therefore be expected to result in the selective loss of K1. Although the toxin or immunity genes of K1 are probably a better initial target for AS inhibition, in that the resultant phenotype should be relatively easy to distinguish, inhibition of K1 Dpol gene expression was considered to be more relevant to this study. Furthermore, initial attempts to inhibit K-plasmid toxin expression, by the synthesis of an AS transcript corresponding to K1 ORF 2, revealed that toxin production was not completely shut off under these circumstances (Romanos, unpublished results). In addition, accurate quantitation of toxin production was difficult.

The mechanisms operative in AS inhibition are only just beginning to be elucidated and, as yet, there is no reliable method for predicting the most susceptible site in an mRNA. Data generated in prokaryotic and viral systems, however, suggest that the most consistent

Figure 7.1 The structure of the *E. coli-K. lactis* shuttle vector used for expression of the antisense RNA in *K. lactis*. This plasmid was derived from a fusion of pG1 and KEp6. The origins of each part of the plasmid are denoted by different internal shading. The antisense gene was inserted between the unique *Bam*H I and *Sal* I sites.



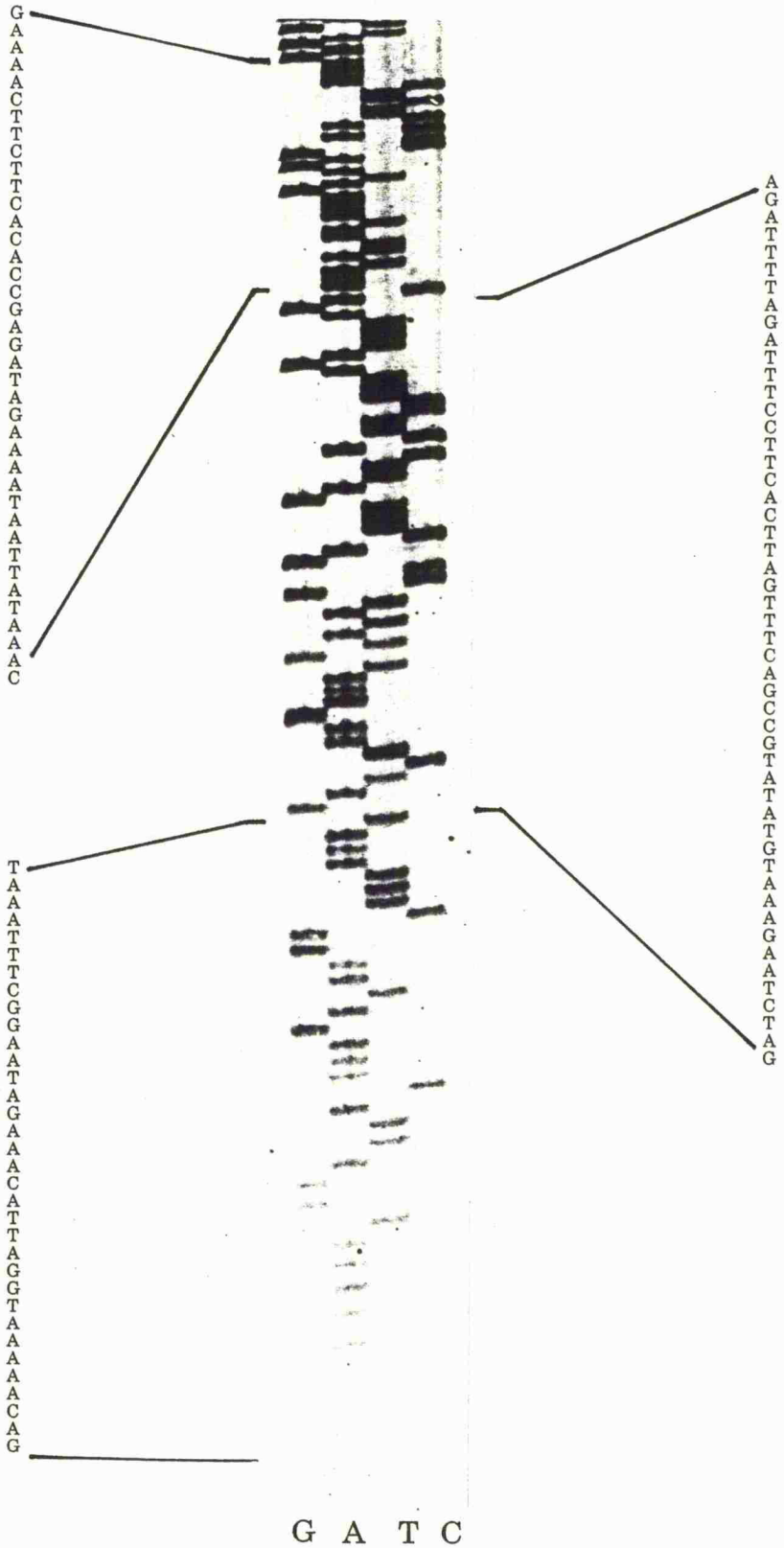
success is obtained with RNAs complementary to the extreme 5' end of an mRNA, including the ribosome binding site and the initiating codon. The transcriptional start site for the K1 Dpol gene has been mapped (Romanos & Boyd, 1988). Accordingly, a 345 bp fragment of the K1 ORF 1, complementary to the 5' untranslated region, the initiating ATG and the first 110 codons (nt 233-575) was selectively amplified using PCR. The two PCR primers CA11 and 12 (Appendix 1) were designed such that the gene fragment could be inserted into the *Bam*H I - *Sal* I sites of the *K. lactis* expression vector pCA15, in the reverse orientation. Transcription of this gene should result in the synthesis of high levels of an AS RNA.

The PCR generated fragment was initially cloned using the vector pCR1000 (Mead *et al.*, 1991). This plasmid was designed to facilitate the direct cloning of PCR fragments, on the basis that some thermostable Dpols used in the PCR catalyse the non-specific addition of an adenine nt to the 3' end of the DNA strand. Cleavage of pCR1000 with *Hph* I yields cohesive ends with a single 3' protruding terminal thymine nt, with which the PCR products can be directly cloned. The structure of plasmid pCR1000 facilitates DNA sequencing of the PCR generated product and the identification of recombinant plasmids bearing the PCR fragment by the use of insertional inactivation of the *lacZ* product. In this case (in accordance with previous observations) the insertion of small DNA fragments into pCR1000 still allowed the cleavage of the chromogenic substrate in the selective media, due to translational 'read-through'. Restriction analysis of plasmids prepared from selected white or light blue colonies revealed the presence of the PCR fragment, and this was confirmed by DNA sequencing (Figure 7.2). This plasmid was termed pCA16. The PCR amplified K1 ORF1 fragment was then excised from pCA16 by a *Bam*H I-*Sal* I digestion, gel purified and cloned into the *Bam*H I-*Sal* I site of pCA15, yielding plasmid pCA17.

7.2.3 Expression of the K-plasmid Specific Antisense Gene.

Plasmid pCA17 was used to transform *K. lactis* JK1 and TRP^+ transformants were isolated on selective media, after about 10-14 days incubation. Examination of the nuclear plasmid

Figure 7.2 Part of the nucleotide sequence of the K1 ORF 1 antisense gene fragment in pCA16.



content of these yeast strains, by plasmid rescue into *E. coli*, and subsequent restriction analysis, confirmed the presence of the plasmid pCA17.

The ability of pCA17 to express the AS RNA was examined by Northern analysis of total yeast RNA. When probed with the radiolabelled 345 bp AS gene fragment, a transcript of approximately 500 nt was detected (Figure 7.3). This transcript was easily detectable and must have been present far in excess of the native K1 ORF 1 transcript, which was not detected by Northern blots (in agreement with Stark *et al.*, 1984; Romanos & Boyd, 1988). The linear plasmid profile of *K. lactis* TRP^+ strains containing the plasmid pCA17, however, revealed that both K1 and K2 were present. Partial inhibition of ORF 1 expression might have led to a reduction in K1 copy number, but this was not obvious in any clone from examination of the plasmids on *in situ* lysis gels. These results suggest that although the AS RNA is being synthesised, it is not inhibiting K1 ORF 1 expression. The alternative is that ORF 1 is not essential for K1 replication.

7.3. DISCUSSION.

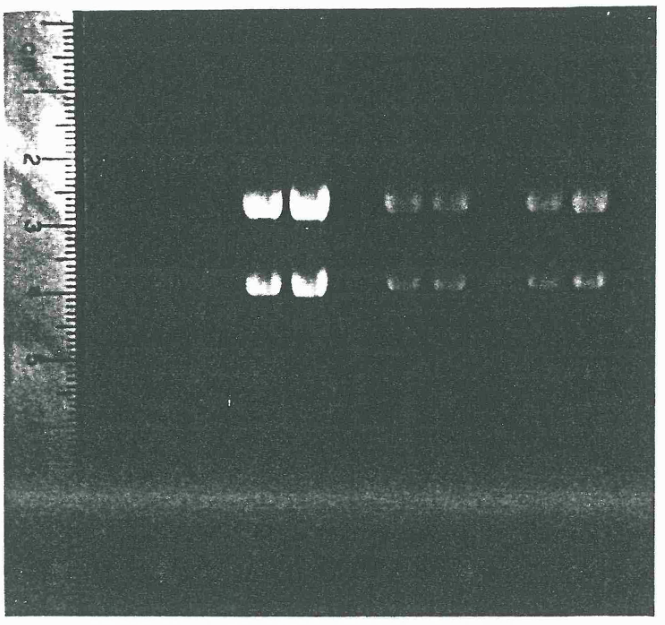
The current literature reveals that AS technology is already proving a valuable tool in the specific modulation of gene expression. Despite this, the mechanisms of AS mediated inhibition, and the reasons why some RNAs are better than others at inhibiting gene expression, are not clearly understood.

In this study, a *K. lactis* nuclear circular vector which directs the high level synthesis of an AS transcript corresponding to the extreme 5' region of ORF 1 of K1, has been constructed. Although high levels of the AS transcript were detected, there did not appear to be any inhibition of K1 ORF 1 expression.

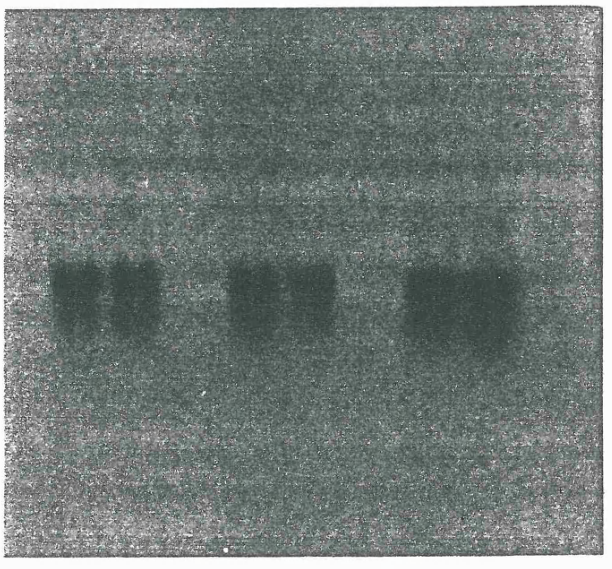
The interaction of sense and AS RNAs is probably governed by a number of factors including the degradative stability of the AS transcript, the existence of secondary structure in either or both the molecules, the stability of such structuring, the relative concentrations and the cellular localisation of the interacting RNAs. The stability of transcripts is thought to be an important control point in the regulation of gene expression. Most RNA molecules do not exist as random coils but form specific structures due to

Figure 7.3 Analysis of *K. lactis* TRP prototrophic transformants carrying the plasmid pCA17, which expresses an antisense RNA to K2 ORF 2. **Panel A.** Duplicate loadings of total *K. lactis* RNA prepared from three independent TRP transformant clones, electrophoresed on a 1.5% w/v denaturing agarose gel. **Panel B.** Northern blot of the gel shown in Panel A, probed with a radiolabelled 5' fragment of K2 ORF 2.

Panel A.



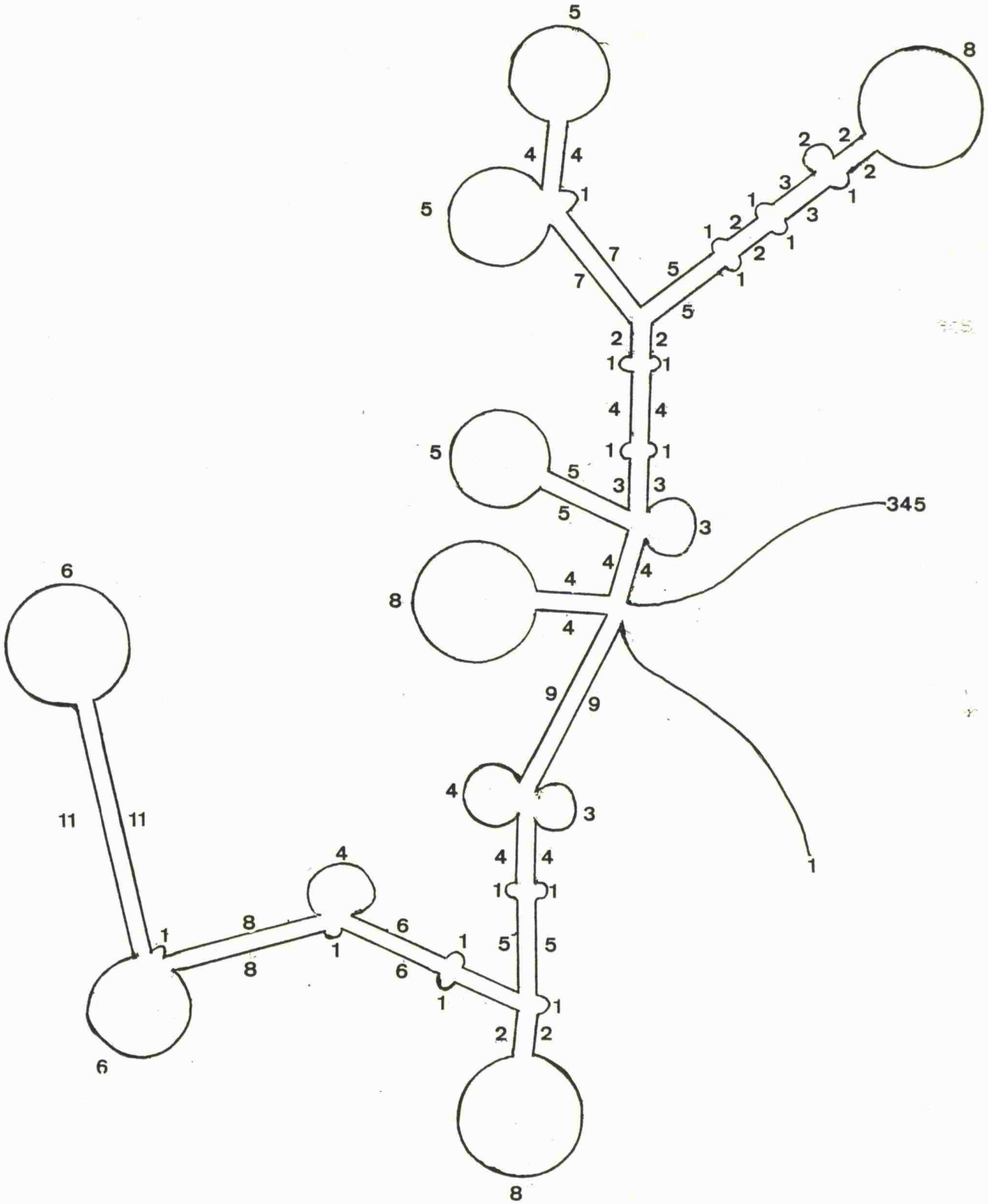
Panel B.



intramolecular hydrogen bonding. Such folding often influences the susceptibility of these molecules to degradation by RNases. The specificity of interaction between two RNA molecules in an extended linear form is determined only by the nt sequence. If, however, one or both of the species are folded, this generates a much greater interactive specificity, and therefore contributes to the exquisite selectivity of AS inhibition. This has been studied most extensively in the ColE1 copy number control system (Euguchi *et al.*, 1991). A small change in the primary structure can, thus, have a profound effect upon the selectivity and rate of such interactions, as illustrated with the IS10 transposase (Kittle & Klechner, cited in Green *et al.*, 1986). Mutations that reduce the interactive capacity of the AS and mRNA result in an increased transposition frequency. The interaction of the two RNA species probably also has to occur during transcription of the mRNA, as these transcripts often assume extensive and stable secondary structure as synthesis progresses. Rapidly folding and highly structured mRNAs with high internal stabilities may not be susceptible to AS inhibition at all. This may explain why the *E. coli ompA* gene is much less susceptible to AS inhibition, in contrast to both the *ompC* and *ompF* genes. Despite the fact that 7 different regions of the *ompA* gene were targeted by AS RNAs, none efficiently inhibited its expression (Coleman *et al.*, 1984). The A+U rich nature of K-plasmid transcripts probably results in a high degree of secondary structuring. Analyses of the potential of the K1 ORF 1 AS RNA to assume higher order structuring (using the 'Fold' programme in the University of Wisconsin Molecular Biology package) reveals extensive folding (Figure 7.4). Furthermore, in view of the high A+U content of the AS RNA, the presence of the polyadenylation sequence at the 3' end of the transcript could conceivably interfere with the ability of the AS RNA to interact with the mRNA.

In some cases hybridisation in the nucleus may prevent transportation to the cytoplasm, or render the hybrid very labile in the cytoplasm. However, as the K-plasmids probably reside and are transcribed in the cytoplasm, this is unlikely to be the problem. Antisense inhibition has also been observed to be dose dependent, thus a high ratio of AS to sense RNA is generally thought be required for efficient inhibition. This has been illustrated with the *E. coli lpp* gene (Coleman *et al.*, 1984), IS10 transposition (Simons & Klechner, 1983) and the *Drosophila kruppel* gene (Rosenberg *et al.*, 1988). As the AS transcripts were easily

Figure 7.4. A potential secondary structure for the small K1 ORF 1 antisense RNA, produced using the University of Wisconsin Molecular Biology package. This programme predicts the folding of short RNA molecules, such that the most stable structure (that with a minimum free energy) is assumed. The figures denote the number of nt involved in the corresponding stems and loops.



detected by Northern analyses of total *K. lactis* RNA, in contrast to the native K1 ORF 1 transcript, a large molar excess of the AS species must exist, precluding the possibility that there was insufficient AS RNA to mediate efficient inhibition.

Antisense inhibition seems to act *via* two main mechanisms. In the first, the AS RNA directly blocks the function of the region to which it binds. This is illustrated by the sequestering of the translational initiating region into a duplex structure, thereby preventing translational initiation (Coleman *et al.*, 1984; Case *et al.*, 1990) or binding directly to the DNA template, preventing transcription (Hoopes & McClure, 1983). The Vesicular Stomatitis virus, by synthesising an AS transcript complementary to the terminal replication origin, is thought to use this mechanism to shut off adenovirus DNA replication in infected cells. This effectively blocks the formation of the initiation complex, by preventing the binding of the pTP (Remenick *et al.*, 1988). In the second mechanism, the AS transcript may function indirectly, by altering the structure of the target to which it binds. This could prevent subsequent processing of the transcript (eg the ColE1 primer), inhibit translation by altering the secondary structure of the messenger and rendering the ribosome binding site or the initiating codon inaccessible, by causing premature transcriptional termination due to the formation of terminator type structures (eg. as with the *E. coli crp* gene), or by destabilisation of the hybrid (Coleman *et al.*, 1984; Robert *et al.*, 1989).

In general, the most effective AS RNAs in prokaryotic and viral systems are those corresponding to the 5' end of the gene, most notably complementary to the upstream untranslated region including the ribosome binding site and the initiating codon (Coleman *et al.*, 1984, Izant & Weintraub, 1984; Kim & Wold, 1985; Melton, 1985; Sczakiel & Pawlita, 1991). This is, however, not always the case and this distinction also appears to be much less clear cut in eukaryotes. Antisense RNAs directed against internal fragments of the *E. coli lacZ* gene have been shown to efficiently inhibit β-galactosidase expression (Pestka *et al.*, 1984; Ellison *et al.*, 1985). The micro-injection, into mouse cells, of AS fragments corresponding to the 3' end of the *CAT* gene, blocked synthesis, although the efficiency was not as great as that observed with 5' sections of the gene (Izant & Weintraub, 1985; Harland & Weintraub, 1985). In transgenic plants, however, an AS RNA complementary to the 5' 172 nt of the CAT gene

was found to be less effective than the whole sequence. On the other hand, an AS RNA corresponding to the 3' half of the nopaline synthase gene was more effective in reducing nopaline synthase activity than a transcript corresponding to the complete gene (Sandler *et al.*, 1988). The general lower effectiveness of AS RNAs corresponding to internal or 3' regions may simply reflect the separation of the hybridised strands by translocating ribosomes. Smaller AS segments may also be more efficient inhibitors as they probably hybridise to the target sequence more rapidly than longer fragments. Nevertheless, long AS transcripts could have certain advantages, in that they may be capable of tolerating the base changes which appear frequently in rapidly evolving viral strains (eg. HIV).

Any of the above reasons could account for the apparent inability of the K1 ORF 1 AS transcript to inhibit expression of this gene. Nevertheless, now that a suitable expression vector is available, a series of AS fragments, corresponding to different regions of the gene, could be constructed and their effects analysed. If a co-ordinated K-plasmid AS programme was to be initiated, a more suitable target, in the first instance, might be ORF 3 of K1. The expression and efficacy of AS transcript production and activity could then be monitored by inhibition of the immunity phenotype. Furthermore, since an *in vivo* homologous recombination system has now been established for the manipulation of K-plasmid genes *in situ*, this could provide a viable alternative to expressing AS RNAs from nuclear vectors. Such AS genes could be inserted into non-essential regions of the K-plasmids, eg. ORF 1 of K2. This might circumvent problems which arise due to transcriptional termination in certain K-plasmid genes. The synthesis of AS transcripts in the immediate vicinity of the native mRNAs may also preclude several of the other problems.

CHAPTER 8

GENERAL DISCUSSION.

Although initially thought to be something of an oddity, the ubiquity of linear plasmids is now becoming increasingly clear. A rapidly escalating number of these extrachromosomal elements is being described, and these seem to be especially prevalent in fungi (see Table 1.4). This thesis has concentrated on the two cytoplasmic linear plasmids which confer a killer and immunity phenotype upon the fungal host, *Kluyveromyces lactis*. The evidence to date, much of it circumstantial, indicates that these episomes belong to a structurally homogenous family of elements, known as Invertrons (Meinhardt *et al.*, 1990). Detailed experimental analyses of several members of this family, primarily Ad, ø29 and PRD1, suggest that the majority of such episomes are self replicating. Both K1 and K2 share the structural features common to this group. Perhaps the three most important features, the presence of ITRs and TPs, and the capacity to encode their own Dpols, relate to the mechanism employed by these Invertrons to replicate their genome. There is as yet, however, a paucity of biochemical evidence to confirm that such a mechanism operates in the K-plasmids, although this has been analysed in considerable detail in Ad and ø29 (reviewed by Salas, 1991).

Therefore in order to provide such evidence, attempts were made to isolate some of the K-plasmid replicatory proteins. The most convenient way of achieving this goal was to employ recombinant DNA technology. The elucidation of the nt sequences and the subsequent definition of 14 ORFs within the K-plasmids facilitated the cloning of these putative genes. As ORFs encoding Class B Dpols have been identified on both K1 and K2 (Fukuhara , 1987; Jung *et al.*, 1987; Tommasino *et al.*, 1988) these were the primary target. The over-expression of genes in the prokaryote *E. coli,* and the isolation and purification of the products, is relatively commonplace nowadays. However, obtaining these proteins was much more difficult than initially expected.

First attempts to isolate the Dpols were made by expressing the complete genes with the aim of synthesising active native proteins, as has been accomplished with Ad, ø29 and PRD1 Dpols (Blanco & Salas, 1984; Shu *et al.*, 1987; Savilahti *et al.*, 1991). When this failed the genes were coupled to protein A gene fragments, with the aim of not only reducing any

product degradation by the host, but also of facilitating product recovery (Nilsson *et al.*, 1985). However, this too proved unsuccessful (Figure 3.5).

Sequence analysis of K-plasmid DNA had previously revealed an extraordinarily high A+T content (Stark *et al.*, 1984; Hishinuma *et al.*, 1984; Wilson & Meacock, 1988; Tommasino *et al.*, 1988). This had probably arisen as a result of the cytoplasmic localisation of these plasmid, and is characteristic of extranuclear genetic elements. On the bases of observations made with other genes, expressed in both this and other systems, an adverse nt composition could have important implications at two major levels: transcription and translation. The lack of expression observed may therefore have been due to inefficient transcription of K-plasmid DNA, as a result of the fortuitous occurrence of sequence structures resembling transcriptional initiators or terminators. In order to minimise any such problems the powerful *E. coli* T7 Rpol transcription (pET) system was utilised. The significance of this lies in the intrinsic biochemical properties of the viral Rpol, and most importantly on its high inherent processivity. Although this was expected to circumvent any transcriptional problems, attempts to express the complete K2 Dpol gene in this system failed. Analysis of transcripts synthesised *in vitro* subsequently revealed that no full length mRNAs were produced from K2 ORF 2, but two species of 1 and 1.4 kb, corresponding to the extreme 5' end of the gene, were detected (Figure 3.4). Prematurely terminated transcripts of this type probably have a much reduced functional lifetime, a possible lack of structuring at the 3' end rendering these transcripts particularly susceptible to RNases. In view of the fact that T7 Rpol is much more processive than *E. coli* Rpol, this alone could account for the the lack of expression observed with the *E. coli* enzyme.

Since these experiments demonstrated that at least 1 kb of K2 ORF 2 was being transcribed, small fragments of less than 1 kb should have been completely transcribed. Nevertheless, attempts to express the first 390 bp of this gene also failed, despite the fact that transcripts of the appropriate size were detected *in vitro*. These results suggested that difficulties, other than at the transcriptional level, were also being experienced.

Following initial (unsuccessful) experiments with protein A fusion vectors, fragments of the K2 Dpol gene were fused to the 3' end of the gene for the enzyme GST . The parental

GST was consistently produced at very high levels and could be isolated very simply via an affinity matrix. The fusion of Dpol gene fragments, of various sizes, however, yielded dramatically different results. In no case was any recombinant fusion protein detected. In addition, upon induction, several of the strains carrying recombinant plasmids exhibited a severe reduction in culture growth rate, something which had first been observed with the pET-derivative containing the 5' fragment of K2 ORF 2. Furthermore, in every case there was a total loss of the parental GST moiety. That this was not due to out-of-frame translational fusions was demonstrated by the insertion of one gene fragment in the reverse orientation. As expected, the translocating ribosome rapidly encountered a nonsense codon, producing a truncated product more or less equivalent to the parental peptide (Figure 3.8).

These experiments led us to conclude that over-expression of the large native K-plasmid genes was probably not possible. This was reinforced by attempts to express ORF 6 of K2, the putative Rpol, in *E. coli* (Brophy & Meacock unpublished results). Evidence presented here demonstrate that only relatively short sequences of K-plasmid DNA could be transcribed, and that mRNA was either unstable, inefficiently translated or the products were rapidly degraded. Furthermore, even when both transcription and translation were efficiently initiated, the Rpol and ribosomes may have 'stalled' soon after encountering the K-plasmid sequences, the pausing of the translocating ribosomes perhaps rendering the mRNA susceptible to degradation. The interpretation of the available data was rendered more difficult by the success experienced with another K-plasmid gene. Small fragments of the K2 encoded Rpol were expressed as fusions with GST, although these products were almost completely insoluble and were not amenable to affinity purification. None of the recombinant Dpol plasmid bearing strains, however, produced an insoluble product.

Although the basis for these differences in expression could not be established by comparison of the nt and codon content of the Rpol and Dpol fragments, it is perhaps significant that the GST-Rpol expressing cultures did not exhibit the severe reduction in growth rate upon induction, in contrast to the GST-Dpol containing strains. This suggested that the GST-Dpol fusions, if indeed they were synthesised, may have been

more toxic than the GST-Rpol fusions. This was also supported, to some degree, by our inability to clone either the K1 or K2 Dpol genes in the constitutive vector pRIT5, which directs heterologous products, in the form of protein A fusions, to the periplasmic space. If the addition of the Dpol gene sequences destabilised the product, one would simply expect degradation of the product. However, the lack of recombinant plasmids obtained suggested that some degree of Dpol gene expression was occurring but that the product was toxic, possibly by causing the protein to become lodged in the cell membrane.

The only remaining alternative, which would preclude the transcriptional and translational problems inherent in the native K-plasmid sequences, with regards to expression in *E. coli*, was to resynthesise the gene. Not only would this allow 'normalisation' of the nt content, but also optimisation of the codon bias relative to that of *E. coli* genes. The considerable discrepancy between the codon bias exhibited by the K-plasmid genes and that of highly expressed *E. coli* genes is most striking when considering the four rarest *E. coli* triplets, all of which are abundant in the K-plasmid ORFs. A fragment corresponding to the extreme N-terminal of K2 ORF 2 was, therefore, redesigned incorporating restriction enzyme sites to facilitate cloning, sequencing and expression in both the pET and pGEX plasmids, and a near optimal codon composition for expression in *E. coli*. Although these measures were anticipated to solve both the transcriptional and translation problems, upon cloning this fragment into pET3a, no recombinant protein was detected, despite the synthesis of a single discrete mRNA *in vitro* (Figure 4.3). In view of the optimised codon content, and the synthesis of the full length transcript, we suspected that the small (5 kDa) product was probably rapidly 'turned over', like the small peptides somatostatin and β-endorphin (Itakura *et al.*, 1977; Shine *et al.*, 1980). This was surmounted by coupling the Dpol sequence to the GST carrier moiety, and upon induction strains bearing these plasmids accumulated the fusion protein to over 10% of total cellular protein (Figure 4.4). Even so, this protein could not be isolated by affinity chromatography, and examination of cellular protein fractions revealed that the vast majority of the protein was produced in an insoluble form (Figure 4.5), as previously observed with the GST-Rpol fusions.

This success should now be consolidated. Clearly there is sufficient protein to allow the generation of polyclonal antibodies to the Dpol fragment, which can in turn be used to detect and isolate the native enzyme from yeast cell extracts. There are, however, complications. Although the gene was designed such that it could be cloned into pGEX2T, which encodes a product containing a proteolytic cleavage site at the GST-insert junction, the fusion is largely insoluble. Therefore, there are a number of options available. Work could proceed with the insoluble fragment already available, or attempts could be made to minimise the amount synthesised in an insoluble form. This might be achieved by altering the kinetics of recombinant protein synthesis, in the simplest cases by reducing the culture growth temperature (Schein, 1990), or the induction rate (Kiefhaber *et al.*, 1991), both of which have proven to be effective in some cases. Nevertheless, these were not effective with the insoluble GST-Rpol fusion (Brophy & Meacock, unpublished results). Alternatively, the next section of the gene, which has already been designed could be synthesised and coupled to the first fragment. Although there is no guarantee that this too would not be insoluble as a fusion with GST, the larger product might be less susceptible to proteolysis as a non-fusion protein.

If the insoluble product which is now available is produced in the form of inclusion bodies, then these can be isolated relatively simply (by differential centrifugation) and used directly to immunise animals. If not the product may have to be purified from SDS-PAGs. However, as the GST moiety is considerably larger than the Dpol fragment, this could result in only a small proportion of the polyclonal antibodies raised being specific for the Dpol, depending upon the antigenicity of the GST fragment. A better approach would be to cleave off the carrier once the protein has been solublised, and then use just the Dpol segment to generate antibodies. A larger Dpol fragment would also be advantagous from this point of view. The use of this Dpol fragment to raise antibodies has two further facets which have to be considered. The purpose of raising the antibodies would be to allow detection and isolation of the authentic product. This, therefore, depends upon the antibodies subsequently recognising the complete native protein. As the immunogen was originally synthesised not only as a fusion product, but also in an insoluble state, this raises questions as to how closely this isolated fragment resembles the corresponding portion in

the native protein. Resolublisation and removal of the GST moiety may allow the refolding of the peptide into its native conformation, as this fragment is located at the N-terminus of the authentic protein, sites which tend to have a relatively high degree of flexibility. Clearly then, there are advantages in not only attempting to maximise the amount of soluble product, but also in synthesising a larger fragment which may more closely resemble the native protein. In terms of simply detecting the authentic protein, the use of the original solublised fragment may not be a problem as native yeast proteins can be presented in a denatured state (eg. by SDS-PAGE).

Even assuming that antibodies are raised which specifically recognise the over-expressed Dpol fragment with or without the GST carrier, the native gene product may not be detected in yeast extracts, possibly because of its low abundance. If so, this may necessitate preparation and concentration of large volumes of cytoplasmic yeast extracts. A reliable large scale procedure which leaves the nuclei largely intact (initially devised to enable large scale isolation of the K-plasmids themselves) has been developed in our laboratory (Reay, Pers. Comm.). This problem may have been experienced with the K2 encoded Rpol synthesised as fusions with both GST and β-galactosidase (Brophy & Meacock, Fleming & Meacock, unpublished results). Although the antibodies raised to the complete fusion products recognised the denatured fusion proteins, those raised to a β-galactosidase fusion recognising a GST fusion (Brophy, Pers. Comm.) neither recognised anything in the yeast extracts.

Because of the difficulties experienced during attempts to express the K-plasmid genes in *E. coli*, other methods of obtaining the Dpol gene products were also considered. To this end three further alternatives were identified, two of which were examined. The first involved the development of an expression system more 'suited' to the extreme nt and codon composition of the K-plasmid genes. A good example is afforded by *S. aureus*. The main limitation here, however, was the unavailability of a regulated *S. aureus* promoter. The use of the *E. coli-S. aureus* shuttle vector pRIT5 (see Chapter 3) was limited by the fact that it directs heterologous protein synthesis in a constitutive manner in *E. coli* (although this might not be a problem in *S. aureus*). This was probably the reason why *E.*

coli transformants containing the K-plasmid Dpol genes could not be isolated. If a regulated expression system did become available, this is more likely to prove successful than any of the other systems currently available.

The second option was to use synthetic peptides to raise antibodies which specifically recognise the conformation of the native protein. This offers the opportunity to raise antibodies to a gene product, directly from the information provided by its nt sequence, without first having to isolate the protein. Unforunately, the application of SPs is currently limited by the inability to locate such sequences in the primary structure. For this reason, the best practical approach is to use all the available methods to analyse the required sequence, and to targets those sites predicted, by the majority of models, to be antigenic. On this basis two sites within the extreme 5' region of the gene were selected as candidates for antigenic peptides. These were synthesised in the form of multiple antigen peptides (Tam, 1988), which in theory dispense with the need for a conjugant protein carrier. However, immunisation of chickens with these peptides did not induce an immune response, probably reflecting the mode of presentation of the peptides rather than their sequence *per se*. It is becoming increasingly clear that these MAP cores are troublesome and that, by far, the best approach is to synthesise linear peptides and couple these to carrier proteins in the conventional manner (Moss, Pers. Comm.). Synthetic peptide technology is currently recognised as being fairly unpredictable in terms of the antibodies raised recognising the complete protein. However, if this work was to continue it would be necessary to resynthesise these peptides in the conventional way and to then immunise both chickens and rodents. In addition I would recommend synthesis of longer peptides (of 10-15 residues) corresponding to both the N- and C-termini, sites which are often good targets.

The third, and final, alternative which was investigated was the possibility of using the K-plasmids themselves to allow detection and isolation of their gene products. There were two ways of doing this. Firstly, by inducing them to over-express their own genes. This would probably involve characterisation of the endogenous promotors, possibly by linking these to reporter genes and reintroducing them into the K-plasmids. Over-expression may,

however, remain problematic due to the limitations imposed by the K-plasmid codon bias or toxicity of the products. The second approach was based upon the modification and reintroduction of K-plasmid genes to facilitate the detection and isolation of the products. Such sequences are reintroduced via homologous recombination between the native plasmids resident in *K. lactis* killer strains and linear K-plasmid derived sequences containing a selectable marker (see Chapter 5). Modified versions of ORFs 4, 5 and 6 of K2 and ORF 1 of K1 have all been reintroduced into K1 in this way. Modifications were made in the form of either a C-terminal antibody binding epitope, 9E10 (Evan *et al.*, 1985) or a metal chelating polyhistidine (Hoculi *et al.*, 1988) 'tag'. The modified versions of the genes have to be segregated from the native untagged versions of the genes, and this is accomplished using the selectable markers. The ability to do so indicates that the tagged proteins are functional. A non-functional gene product would probably be reflected by a loss of the plasmid, assuming the product was essential for replication. Such a situation is, therefore, characterised by an inability to segregate the native and recombinant plasmids.

In this study the putative Dpol gene of K1 was modified by the addition of a C-terminal polyhistidine tag. This was then reintroduced into K1, not into ORF 2 as with the other genes, but directly back into ORF1. The integration event was projected to result in the generation of a single plasmid, slightly longer than the native K1 species. However, transformation yielded a variety of recombinant K1 derivatives, some slightly larger, and others slightly smaller than native K1. Attempts to segregate the native and recombinant K1 plasmids yielded some interesting results. Although very little, if any, native K1 remained, two of the three *K. lactis* strains examined retained two K1-derived plasmids. The third strain retained only a single version of K1, shorter than the native K1. These results, therefore, suggested that an unexpected recombination event had taken place. This, presumably, arose as a consequence of the structure of the linear integrating fragment, which contained a 1 kb duplication of the 3' end of K1 ORF 1. In all probability the linear termini were responsible for the initial recombination, but the presence of a sequence duplication resulted in subsequent plasmid rearrangement and deletion of the intervening sequence. This is supported by the fact that in the strains retaining two versions of K1, both appear to carry the selectable marker. This event may also be

influenced by the possibility that the tagged version of the ORF 1 product is non-functional; rearrangement and deletion of the tagged termini allowing the reformation of a complete, functional gene. These results need to be confirmed and in addition, which of the K1 derivatives retains the tagged version of ORF 1 must be determined. Depending upon the results of these experiments, an attempt could be made to isolate the ORF 1 product. That this technique holds great potential to solve some of the problems that have been experienced was demonstrated by the identification of the K2 ORF 5 product (Schaffrath, unpublished results). Furthermore, the utility of both these tags has also been demonstrated using C-terminal fusions with by K2 ORF 7 expressed in *E. coli* using the T7 Rpol system (Schaffrath, unpublished results).

The initial goal, to isolate the Dpols, has therefore not been achieved. Progress towards the establishment of a minimal *in vitro* replication system has been delayed by the inability to express the complete K-plasmid genes in *E. coli*. This has been shown to be due to both inefficient transcription and translation. For similar reasons, expression of the complete genes in *S. cerevisiae* would probably have been unsuccessful. It may be significant that the expression of the Dpol genes from other non-nuclear linear plasmids has also not been reported. Their high A+T content makes it extremely likely that expression of these genes would also be dogged by problems of a similar nature to those reported here for the K-plasmid genes.

However, a fragment of the K2 Dpol has now been expressed to very high levels. This success should be consolidated, by the generation of polyclonal antibodies to the Dpol fragment, with the aim of detecting and isolating the native enzyme from yeast cell extracts. In addition, the homologous recombination technology has shown great potential to yield the native enzymes, something which, in all eventualities, may not be feasible via other routes. Despite this, the *in vitro* replication system for the K-plasmids remains a very distant possibility. The antisense technology has only a limited applicability for analysing K-plamid gene function, and further analyses I would suggest should be postponed. Much the same applies for the synthetic peptide approach. Although this may

eventually allow specific antibodies which recognise the native protein to be raised, this is expensive and the results are unpredictable.

CHAPTER 9.

ERRATUM.

9.1 Expression in *E. coli*.

9.1.1 Expression of K2 ORF 2 as a non-fusion protein.

9.1.2 Expression of K1 ORF 1 and K2 ORF2 as fusion proteins.

9.2 Manipulation *in vivo* of the linear plasmids.

Since this Thesis was originally submitted, a number of further experiments have been performed which have clarified a number of points. These are dealt with here, in the order with which they occur in the previous Chapters.

9.1 EXPRESSION IN *E. COLI.*

9.1.1 Expression of K2 ORF 2 as a non-fusion protein.

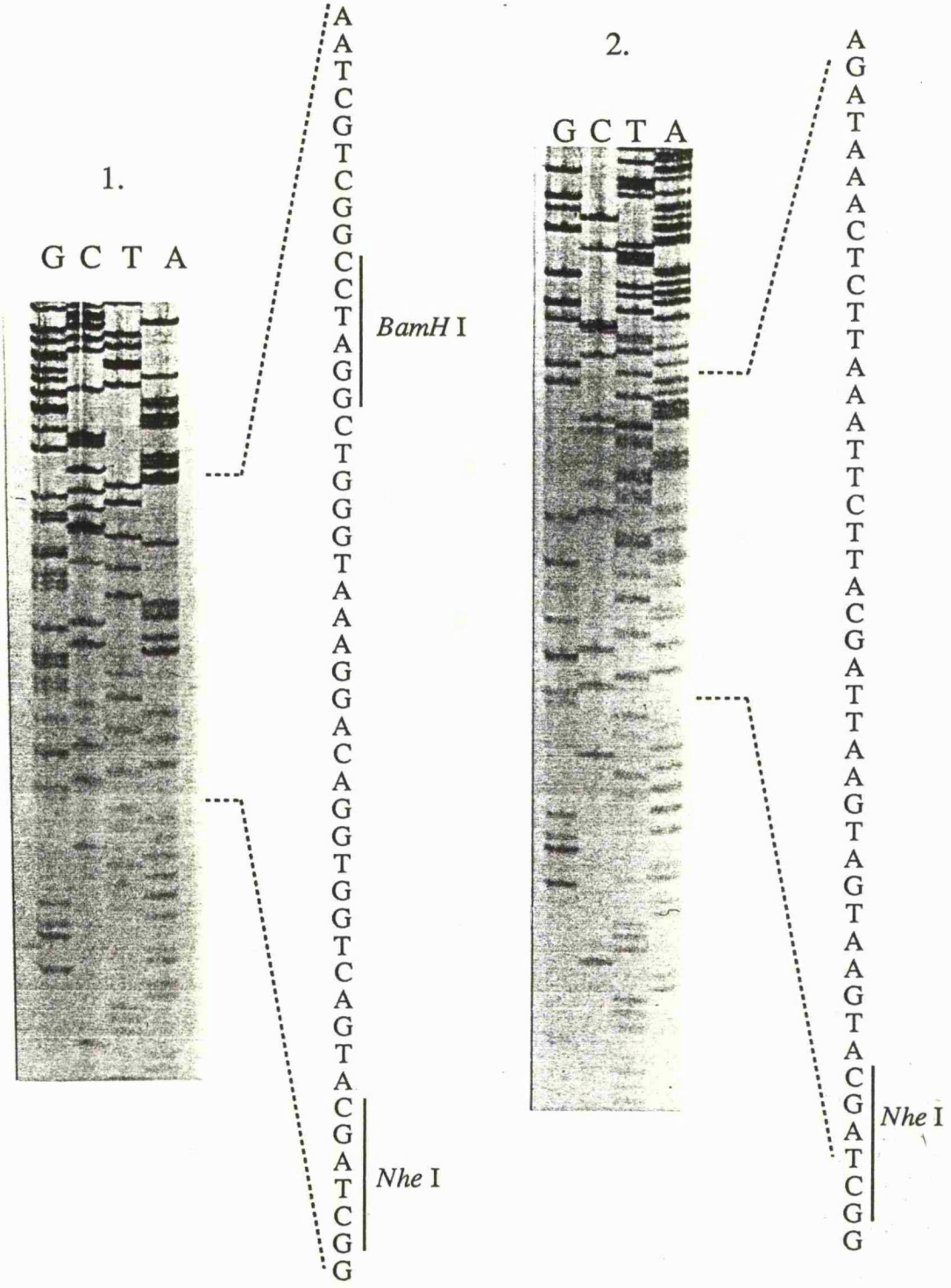
The majority of the possible causes concerning the inability to express the native K2 Dpol gene (ORF 2) as either the entire protein or small fragments of the protein have been investigated. In each case multiple independent DNA clones were analysed by restriction enzyme digestion, and multiple bacterial transformants were examined for their ability to express the heterologous protein. These confirmed that the lack of product was not due to heterogeneity amongst the DNA constructions or the bacterial host. The ability to induce truncated GST-Dpol fusion proteins in the pGEX vectors by constructing out of frame translational fusions (pCA14) suggested that the lack of expression observed with recombinant plasmids pCA10, 11 and 12 was not simply due to cloning into the wrong vector. This was confirmed by the successful expression of a synthetic K2 ORF2 gene fragment in pGEX2T (pCA30) and with K2 Rpol-GST fusions (Brophy & Meacock, unpublished results); the source of the pGEX plasmids in all these cases being identical. Previous experiments with plasmids pCA9 and pCA13, derivatives of the pET vectors containing the complete 3 kb gene (pCA9) and the extreme 5' 390 bp (pCA13), had also demonstrated that T7 Rpol was not capable of efficiently transcribing the entire gene. However, transcripts corresponding to the 390 bp 5' fragment were detected. In both cases no protein product was observed, despite extensive investigation. Furthermore, when the extreme 160 bp of the gene was resynthesised, incorporating an optimal codon bias to facilitate high level expression in *E. coli*, and inserted into the unmodified pET3a vector (pCA29), protein production was still not detected, even though coupling of the same gene fragment to GST (pCA30) resulted in very high levels of the recombinant protein being synthesised.

The vectors pET3a , b and c had been prepared simultaneously in our laboratory, so there was the possibility that the plasmid stocks had been mislabelled. This could mean that the

K2 ORF 2 gene fragments had effectively been cloned translationally out-of-frame in the expression vector. With regards to the recombinant plasmids pCA9 and pCA13, it was irrelevant into which of the three vectors (pET3a, b or c) the gene was cloned, in terms of gene expression. The reason for this was that the reading frame alterations in the pET3 vector only affect the *Bam*H I cloning site about 11 codons downstream of the initiating codon (Figure 3.2). In pCA9 and 13, the ORF 2 gene fragments were not simply cloned into the *Bam*H I site. The pET3a plasmid had previously been modified by an *Eco*R V deletion (pCA8, Figure 3.2). This rendered the *Nhe* I site immediately downstream of the initiating ATG in pCA8 unique. The PCR generated ORF 2 fragments were therefore cloned into *Nhe* I - *Bam*H I cleaved pCA8. As the *Nhe* I site is upstream of the reading frame variations at the *Bam*H I site, the reading frame at the *Nhe* I site is identical in all three vectors. Therefore the ORF 2 fragments should all have been translationally in frame. This has been confirmed by a DNA sequence analysis of the chimeric junction in pCA9 (Figure 9.1).

The inability of the pET3a-synthetic gene plasmid (pCA29) to direct the synthesis of a protein product was perhaps the most worrying of these results. The construction of this plasmid was, however, slightly different from pCA8 and 13. Here, the synthetic gene was sub-cloned from pCA28 as a *Bam*H I-*Bgl* II fragment into the *Bam*H I site of the unmodified pET3a. In this case the reading frame around the cloning *Bam*H I site would directly influence the nature of the translational fusion. The stock of the pET3a plasmid used for this construction was identical to that used earlier for the construction of pCA8. Therefore, although the use of pET3a, b or c would not have affected the expression of the native gene constructions, in this case only pET3a would be expected to yield a protein product. Sequence analysis of the pET3a stock has subsequently demonstrated that this plasmid stock is actually pET3c. This, in all probability, now explains the lack of expression observed with the synthetic gene in this vector, although proteolysis of the relatively small (5 kDa) protein may also influence the absolute amount of product obtained. In addition, this has important repercussions with respect to the further progress of this project; synthesis of the protein in a soluble non-fusion form would greatly

Figure 9.1 DNA sequence analyses of the plasmid stock pET3a (Panel 1) and recombinant plasmid pCA9 (panel 2). The pET3a stock was used for the construction of the *EcoR* V deletion plasmid pCA8, for the native K2 ORF 2 expression plasmids pCA9, 13 and 29. Sequence 1 demonstrated that the plasmid presumed to be pET3a was actually pET3c; this explained the lack of expression obtained with plasmid pCA29. Sequence 2 demonstrates that the native K2 ORF 2 was cloned in frame into the *Nhe* I site of plasmid pCA8.



facilitate the generation of polyclonal antibodies, circumventing the problems discussed in Section 4.3.

9.1.2 Expression as a fusion protein.

The utility of IgG sepharose affinity chromatography to efficiently isolate *Staphylococcal* protein A products, whether in the form of the non-fusion control 28 kDa peptide or the much larger protein A-ORF 1 or 2 fusions, has been called into question by our inability to isolate significant quantities of these products (Section 3.2.2.1b). With respect to the fusion proteins, this could be due in part to reduced product solubility, as has been seen with the GST fusions (see Chapters 3 and 4).

An alternative approach for identifying any protein A fusion products is to utilise the ability of the protein A moiety to bind IgG molecules in a Western immunoblot type analysis. This can be performed simply by electrophoresing total protein extracts from induced and uninduced recombinant plasmid-bearing strains, transferring the proteins to a nitrocellulose membrane (Section 2.6.3) and then probing the membrane with IgG molecules conjugated to a reporter enzyme. This approach has been used to demonstrate the expression of the *K. lactis* α-toxin subunit in *E. coli* as a protein A fusion (Stark, Pers. Comm.). Nevertheless, only approximately 10% of the protein A-toxin product was full length, the remainder consisting of prematurely terminated products. In this case cell extracts prepared from induced (for 3 h) and uninduced *E. coli* N4830-1 strains bearing plasmids pCA3 and pCA4, and the parental plasmid pRIT2T were electrophoresed on a 12% w/v SDS-polyacrylamide gel. No protein A or protein A-fusion products were immediately obvious by coomassie staining of a duplicate gel. The proteins were then transferred to a nitrocellulose membrane by electro-blotting and treated as outlined in Section 2.6.3. Both a primary and secondary antibody (both provided by S. Saville, Department of Genetics, Leicester University) were used to increase the sensitivity of product detection; a pre-immune rabbit sera was used as the primary antibody source, and the secondary antibody was a goat anti-rabbit IgG with a peroxidase conjugant. In order to minimise any background colour development, the primary antisera was used at a dilution of 1 in 1500, and the secondary antibody at a dilution of 1 in 3000. The results are shown

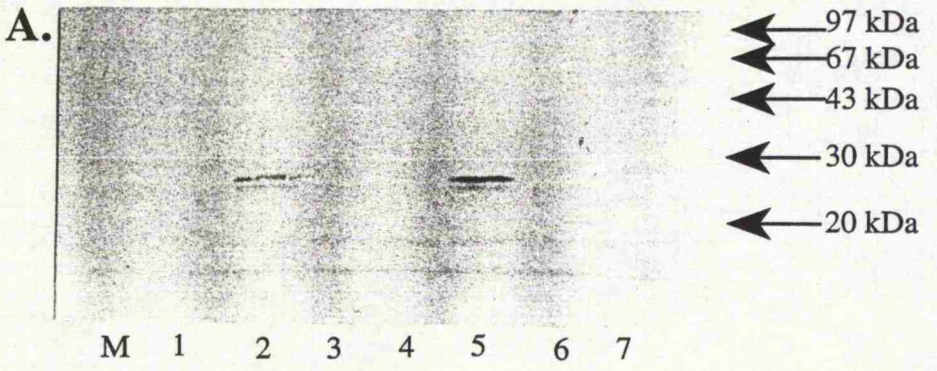
in Figure 9.2. Very little signal was picked up from the uninduced control extracts, but in those containing the parental pRIT2T plasmid the 28 kDa protein A molecule was visible. This was the strongest band observed (Figure 9.2A and B). Further development of the filters revealed different protein profiles between strains bearing pRIT2T, pCA3 and pCA4. The parental protein A product was absent in strains bearing pCA3 and 4, however, additional products were visible in these protein extracts. The strain containing pCA3 produced minor proteins of approximately 40 and 60 kDa, whilst pCA4 bearing strains contained a product of approximately 50 kDa. All these products including the 28 kDa parental protein A molecule, appeared to be of relatively low abundance, as evidenced by the rather weak colour reaction observed. This may have been due to poor heat induction of heterologous protein synthesis or simply due to insufficient time for the heterologous products to accumulate as the *E. coli* host strain (N4830-1) grows relatively slowly.

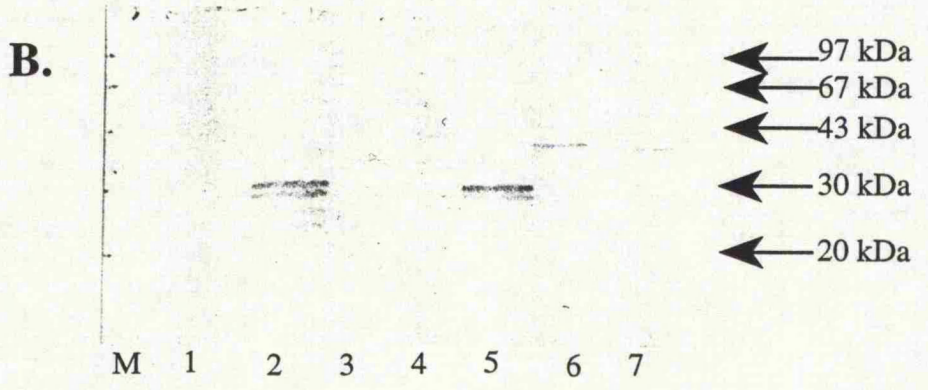
These results demonstrate that protein A-Dpol fusion products were actually being synthesised, but in a very heterogeneous manner and at low concentrations. This also suggests the possibility of using the same approach, should the appropriate antibodies be available, for detecting GST-Dpol fusion products synthesised from any of the recombinant plasmids.

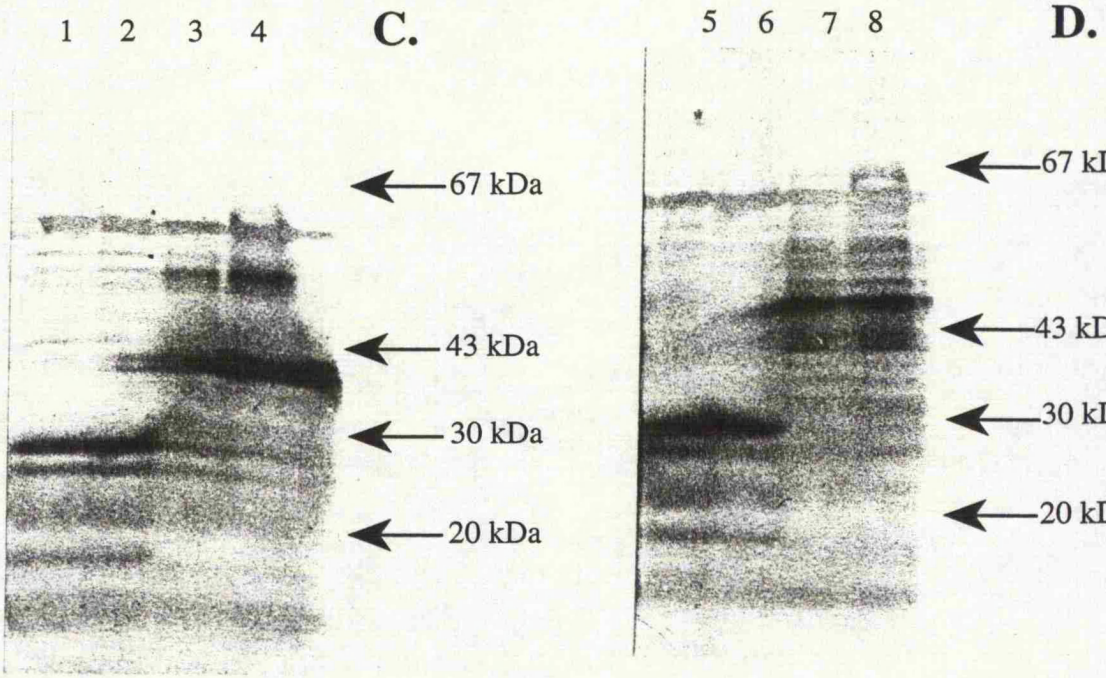
9.2 MANIPULATION *IN VIVO* OF THE LINEAR PLASMIDS.

At the time this Thesis was initially submitted the exact structure of the recombinant K1 plasmids which had become established in *K. lactis* JK1 was unclear. We knew that the integration of the transforming DNA fragment by homologous recombination had occurred as expected, but what we had not expected was the heterogeneity of the recombinant plasmids produced. Southern analyses of the yeast colonies bearing these plasmids immediately after the initial transformation revealed that all contained the *TRP1* gene (Figure 5.5B). Prolonged subculturing of three such strains (1, 2 and 4) under selective conditions led to the loss of the native K1 plasmid, as demonstrated by Southern analyses using a probe derived from part of K1 ORF2 deleted by the initial homologous recombination event (Figure 5.5D). After 120 generations of growth in selective media

Figure 9.2 Western analyses of the *Staphylococcal* protein A-Dpol fusion proteins produced from recombinant plasmids pCA3 and pCA4 in *E. coli* N4830-1. Total protein from these strains was analysed by 12.5% SDS-PAGE, using the Hoeffer mini-gel system, and the protein transfered to a nitrocellulose membrane as outlined in Chapter 2. Both plasmids contain the complete 3 kb ORFs corresponding to the putative K-plasmid Dpols; pCA3 contains K1 ORF 1, pCA4 K2 ORF 2. **Panels A and B.** These membranes differ only in the length of time that the colour reaction was allowed to develop. Lanes M, protein molecular weight markers; Lanes 1, uninduced strain containing the parental plasmid pRIT2T; Lanes 2 and 5, induced strain containing pRIT2T; Lanes 3 and 4, induced strain containing pCA3; Lanes 6 and 7, induced strain containing plasmid pCA4. **Panels C and D.** Western analysis of a greater volume of the same protein extracts, using a larger (40 ml) 12.5% SDS-PAGE. Lanes 1, 2, 5 and 6, induced pRIT2T; Lanes 3 and 4, induced pCA3; Lanes 7 and 8, induced pCA4.





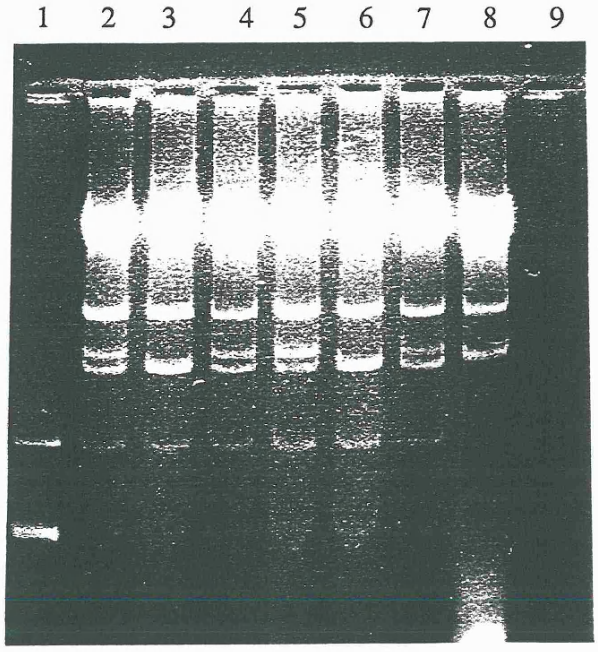


two of the three strains examined (2 and 4) retained two K1 derived plasmids. In these strains, one of the K1-derivatives was slightly smaller than native K1 and the other slightly larger. In the remaining strain (1) which retained only one K1-derived plasmid, this was of approximately the same size as the smaller of the other two K1 derivatives. It may be significant that the abundance of the K1 derivative appeared to be greater when only one species was present in the TRP prototrophic *K. lactis* strain 1; when two K1-derived species were present (in strains 2 and 4), both species appeared to be less abundant. The origin of the smaller plasmid in both strains 2 and 4, we speculated, might lie in the duplication of the 3' region of ORF 1, leading to a structural rearrangement and deletion of the intervening duplicated sequence. Presumably in that strain (1) retaining only a single K1-derived species, the deletion event was subsequently followed by its stable maintenance due to the reformation of a functional ORF 1 product, and loss of the larger precursor. Therefore, in those strains (2 and 4) retaining two K1 derived species, either one of the species does not produce a functional ORF 1 product, or the deletion event is an ongoing phenomenon. This has important consequences with respect to the isolation of the K1 ORF 1 product due to the location of the polyhistidine tag at the 3' end of the outer ORF 1 sequence. Any deletion of this particular sequence would render the isolation of the ORF 1 product impossible. Therefore it was essential to be sure of the exact structure of the recombinant plasmids resident in these strains, and in particular, whether they retained the 'tagged' version of ORF 1. This was accomplished relatively easily by an additional Southern analysis using the radiolabelled polyhistidine tag oligonucleotide. The 27 nt polyhistidine oligonucleotides were end labelled using polynucleotide kinase (Pharmacia) and γ-[^{32}P]-dATP (Maniatis *et al.*, 1982). The probe hybridised strongly to the control plasmid, a pIC vector containing a histidine tagged version of K2 ORF 4 (supplied by Mr. S. Soond, Leicester Biocentre). The hybridisation temperature of the probe was calculated to be between 54^{o}C and 62^{o}C, depending upon the method used for the calculation (Section 2.4.19), so hybridisation and washing was initially performed under conditions of low stringency. After washing at 45^{o}C in 5xSSC, the probe was found to be weakly bound to both K1 and K2 with about equal intensity and also to the *TRP1* control plasmid pMS11. Rewashing the filter under more stringent conditions (3xSSC)

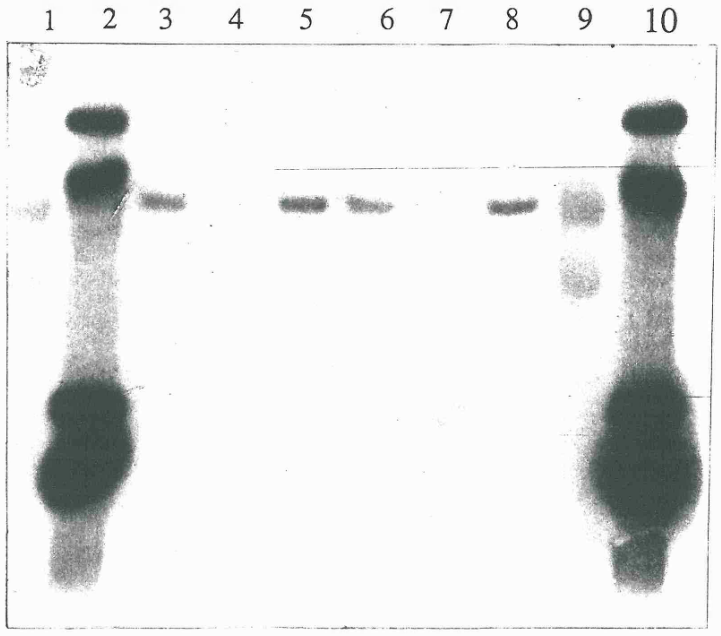
subsequently removed most of the non-specific binding and illustrated that the histidine tag was intact on the larger of the two K1-derivatives (Figure 9.3). The remainder of the non-specific probe hybridisation was removed by increasing the washing temperature to 55^{o}C (data not shown). The signal obtained by autoradiography was, however, quite weak and implies a low copy number. The fact that the smaller deleted K1 derivative is stable in Strain 1, suggests that the ORF1 product in the larger, undeleted species is non-functional, although the possibility exists that a slightly different recombination event originally occurred in this strain.

It may still be possible to isolate the K1 ORF 1 gene product via the polyhistidine tag, but I think that, in view of the low abundance of the tagged gene and that the functionality of the product is unknown, a better approach would be to start again and proceed as discussed in Section 5.3.

Figure 9.3 Southern analysis of the recombinant linear K1-derived plasmids in TRP prototrophic *K. lactis* strains 1, 2 and 4. **Panel A.** A 0.65% w/v agarose *in situ* lysis gel showing the DNA content of the three *K. lactis* strains. Lanes 1 and 9, pIC-ORF4:his uncut circular plasmid (hybridisation control); lanes 2 and 5, recombinant strain 2; lanes 3 and 6, recombinant strain 1; lanes 4 and 7, recombinant strain 4; lane 8, non-recombinant K. lactis JK1 control. **Panel B.** A Southern blot of a similar gel (lacking only the non-recombinant *K. lactis* JK1 strain). Lanes 1 and 9, pMS11 *TRP1* uncut circular plasmid (hybridisation control); lanes 2 and 10, pIC-ORF4:his uncut circular plasmid (hybridisation control); lanes 3 and 6, recombinant strain 2; lanes 4 and 7, recombinant strain 1; lanes 5 and 8, recombinant strain 4.



A.



B.

CHAPTER 10.

REFERENCES.

Adhya, S., Shneidman, P.S. & Hurwitz, J. (1986) Reconstruction of adenovirus replication origins with a human nuclear factor I binding site. **J. Biol. Chem.** 261 (7) 3339-46

Aiba, H., Matsuyama, S.H., Mizumo, T. & Mizushima, S. (1987) Function of *micF* as an antisense RNA in osmoregulatory expression of the *ompF* gene in *E. coli.* **J. Bacteriol.** 169 (7) 3007-12

Andersen, J., Forst, S.A., Zhao, K., Inouye, M & Delihas, N. (1989) The function of *micF* RNA. **J. Biol. Chem.** 264 (30) 17961-70

Argos, P. (1988) A sequence motif in many polymerases. **Nucleic Acid Res.** 16: 9909-16

Ariga, H., Klein, H., Levine, A.J. & Horwitz, M.S. (1980) A cleavage product of the adenovirus DNA binding protein is active in DNA replication *in vitro.* **Virology** 101: 307-10

Arnon, R. (1989) Peptides as immunogens: Prospects for synthetic vaccines. In **Curr. Top. Micro. Immunol.** 130: 1-12

Arnon, R., Maron, E., Sela, M. & Anfinsen, C.B. (1971) Antibodies reactive with native lysozyme elicited by a completely synthetic antigen. **Proc. Natl. Acad. Sci.** 68 (7) 1450-5

Arrand, J.R. & Roberts, R.J. (1979) The nucleotide sequence at the termini of adenovirus-2 DNA. **J. Mol. Biol.** 128: 577-94

Balbas, P. & Bolivar, F. (1990) Design and construction of expression plasmid vectors in *E. coli.* **Meth. Enzymol.** 185:14-37

Balbinder, E. & Walden, C. (1991) A review of DNA metabolism in *E. coli.* **Cell Biol. Rev.** 25 (2) 105-59

Bamford, D., McGraw, T., MacKenzie, G. & Mindich, L. (1983) Identification of a protein bound to the termini of bacteriophage PRD1 DNA. **J. Virol.** 47 (2): 311-16

Bamford, D.H. & Mindich, L. (1984) Characterisation of the DNA protein complex at the termini of the bacteriophage PRD1 genome. **J. Virol.** 50 (2) 309-15

Bartenschlager, R. & Schaller, H. (1988) The amino-terminal domain of the heptadnaviral P-gene encodes the terminal protein (genome linked protein) believed to prime reverse transciption. **EMBO J.** 7 (13) 4185-92

Bedinger, P., Hostos, E.L., Leon, P. & Walbot, V. (1986) Cloning and characterisation of a linear 2.3 kb mitochondrial plasmid of maize. **Mol. Gen. Genet.** 205:206-12

de Benedetti, A., Joshi, Barve, S., Rinker-Schaeffer, C. & Rhoads, R.E. (1991) Expression of antisense RNA against initiation factor eIF-4E mRNA in HeLa cells results in lengthened cell division times, diminished translation rates and reduced levels of both eIF-4E and the p220 component of eIF-4F. **Mol. Cell. Biol.** 11 (11) 5435-45

van Bergan, B.G.M., van der Ley, P.A., van der Mansfeld, A.D.M. & van derVliet, P.C. (1983) Replication of origin containing adenovirus DNA fragments that do not carry the terminal protein. **Nucleic Acid Res.** 11 (7) 1975-80

van der Berg, J.A., van der Laken, K.J., van Ooyen, A.J.J., Renniers, T.C.H.M., Rietveld, K., Schaap, A., Brake, A.J., Bishop, R.J., Schultz, K., Moyer, D., Richman, M. & Shuster, J.R. (1990) *Kluyveromyces* as a host for heterologous gene expression: Expression and secretion of prochymosin. **Biotechnology** 8: 135-9

Bernad, A., Zaballos, A., Salas, M. & Blanco, L. (1987) Structural and functional relationships between prokaryote and eukaryote DNA polymerases. **EMBO J** 6(13) 4219-25

Bernad, A., Blanco, L. & Lazaro, J.M. (1989) A conserved 3' to 5' exonuclease active site in prokaryote and eukaryote DNA polymerases. **Cell** 59: 219-28

Bernad, A., Blanco, L. & Salas, M. (1990a) Site directed mutagenesis of YCDTDs amino acid motif of Ø29 DNA polymerase. **Gene** 94: 45-51

Bernad, A., Lazaro, J.M., Salas, M. & Blanco, L. (1990b) The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in α-like DNA polymerases is required by phage ø29 DNA polymerase for protein primed initiation and polymerisation. **Proc. Natl. Acad. Sci.** 87: 4610-14

Bertrand, H. & Griffiths, A.J.F. (1989) Linear plasmids that integrate into the mitochondrial DNA in *Neurospora.* **Genome** 31: 155-9

Bertrand, H., Chan, B.S-S. & Griffiths, A.J.F. (1985) Insertion of a foreign nucleotide sequence into mitochondrial DNA causes senescence in *Neurospora intermedia.* **Cell** 41: 877-84

Bertrand, H., Griffiths, A.J.F., Court, D.A. & Cheng, C.K. (1986) An extrachromosomal plasmid is the etiological precursor of KalDNA insertion sequences in the mitochondrial chromosome of senescent *Neurospora.* **Cell** 47:829-37

Birnboim, H.C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. **Nucleic Acid Res.** 7: 1513-23

Bittle, J.L., Houghten, R.A., Alexander, H., Shinnick, T.M., Sutcliffe, J.G., Lerner, R.A., Rowlands, D.J. & Brown, F. (1982) Protection against foot-and-mouth disease by immunisation with a chemically synthesized peptide predicted from the viral nucleotide sequence. **Nature** 298: 30-3

Blackburn, E.H. (1991) Telomeres. **Trends In Biochem. Sci.** 16/10/91 p378-81

Blanco, L. & Salas, M. (1984) Characterisation and purification of a phage ø29 encoded DNA polymerase required for the initiation of replication. **Proc. Natl. Acad. Sci.** 81: 5325-9

Blanco, L. & Salas, M. (1985a) Characterisation of a 3'-5' exonuclease activity in the phage ø29 encoded DNA polymerase. **Nucleic Acid Res.** 13 (4) 1239-49

Blanco, L. & Salas, M. (1985b) Replication of phage ø29 DNA with purified terminal protein and DNA polymerase: Synthesis of full length ø29 DNA. **Proc. Natl. Acad. Sci.** 82:6404-08

Blanco, L. & Salas, M. (1986) Effect of aphidicolin and nucleotide analoges on the phage ø29 DNA polymerase. **Virology** 153: 179-87

Blanco, L., Garcia, J.A., Penelva, M.A. & Salas, M. (1983) Factors involved in the initiation of phage ø29 DNA replication *in vitro*: requirement of the gene 2 product for formation of the protein p3-dAMP complex. **Nucleic Acid Res.** 11 (5) 1309-15

Blanco, L., Garcia, J.A. & Salas, M. (1984) Cloning and expression of gene 2, required for the protein-primed initiation of the *B.subtilis* phage ø29 DNA replication. **Gene** 29: 33-40

Blanco, L., Gutierrez, J., Lazaro, J.M., Bernad, A. & Salas, M. (1986) Replication of phage ø29 *in vitro*: role of the viral protein p6 in initiation and elongation. **Nucleic Acid Res.** 14 (12) 4923-36

Blanco, L., Prieto, I., Gutierrez, J., Bernad, A., Lazaro, J.M., Hermoso, J.M. & Salas, M. (1987) Effect of NH_4^+ ions on ø29 DNA - protein p3 replication: Formation of a complex between the terminal protein and the DNA polymerase. **J. Virol.** 61 (12) 3983-91

Blanco, L., Bernad, A., Lazaro, J.M., Martin, G., Garmendia, C. & Salas, M. (1989) Highly efficient DNA synthesis by the phage ø29 DNA polymerase. **J. Biol. Chem.** 264 (15) 8935-40

Blanco, L., Bernad, A. & Salas, M. (1992a) Evidence favouring the hypothesis of a conserved 3'-5' exonuclease active site in DNA-dependent DNA polymerases. **Gene** 112: 139-44.

Blanco, L., Bernad, A., Esteban, J.A. & Salas, M. (1992b) DNA-independent deoxynucleotidylation of the ø29 terminal protein by the ø29 DNA polymerase. **J. Biol. Chem.** 267 (2) 1225-30

Blasco, M.A., Blanco, L., Pares, E., Salas, M. & Bernad, A. (1990) Structural and functional analysis of temperature sensitive mutants of phage ø29 DNA polymerase. **Nucleic Acid Res.** 18 (16) 4763-70

Blasco, M.A., Bernad, A., Blanco,L. & Salas, M. (1991) Characterisation and mapping of the pyrophosphorolytic activity of the phage ø29 DNA polymerase. **J. Biol. Chem.** 266 (12) 7904-9

Blasco, M.A, Lazaro, J.M., Bernad, A., Blanco, L. & Salas, M. (1992) ø29 DNA polymerase active site. **J. Biol. Chem.** 267 (27) 19427-43

Blum, P, Velligan, M, Lin, N. & Martin, A. (1992) DNA-K mediated alteration in human growth hormone protein inclusion bodies. **Biotechnology** 10: 301-3

de Boer, H.A. & Kastelein, R.A. (1986) Biased codon usage: An exploration of its role in optimisation of translation. **In Maximizing gene expression.** Reznikoff, W. & Gold, L. (Ed) Butterworths, pp 225-85

de Boer,H.A. & Hui, A.S. (1990) Sequences within ribosome binding sites affecting m RNA translatability and methods to direct ribosomes to single mRNAs. **Meth. Enzym.** 185: 103-14

Bolziau, C., Thuong, N.T., & Toulme, J-J. (1992) Mechanisms of the inhibition of reverse transcription by antisense oligonucleotides. **Proc. Natl. Acad. Sci.** 89: 768-72

Bosch, V., Bartenschlager, R., Radziwill, G. & Schaller, H. (1988) The duck hepatitis B virus P-gene codes for a protein strongly associated with the 5' end of the viral DNA minus strand. **Virology** 166: 475-85

Bradford, M.M. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein using the principle of protein-dye binding. **Anal. Biochem.** 72: 248-54

Brawerman, G. (1987) Determinants of mRNA stability. **Cell** 48: 5-6

Brodin, P., Grundstrom, T., Hofmann, T., Drakenberg, T., Thulin, E. & Forsen, S. (1986) Expression of bovine intestinal calcium binding protein from a synthetic gene in *E. coli* and characterisation of the product. **Biochemistry** 25: 5371-7

Carusi, E.A. (1977) Evidence for blocked 5' termini in human adenovirus DNA. **Virology.** 76: 380-94

Case, C.C., Simons, E.L. & Simons, R.W. (1990) The IS10 transposase mRNA is destabilised during antisense control. **EMBO J.** 9 (4) 1259-66

Cesareni, G, Helmer-Cittench, M & Castagnoli, L. (1991) Control of *ColE1* plasmid replication by antisense RNA. **Trends In Genet.** 7(7) 230-5

Challberg, M.D. & Kelly, T.J. (1979a) Adenovirus DNA replication *in vitro*. **Proc. Natl. Acad. Sci.** 76 (2) 655-9

Challberg, M.D. & Kelly, T.J. (1979b) Adenovirus DNA replication *in vitro*: Origin and direction of daughter strand synthesis. **J. Mol. Biol.** 135: 999-1012

Challberg, M.D. & Kelly, T.J. (1981) Processing of the adenovirus terminal protein. **J. Virol.** 38 (1) 272-7

Challberg, M.D. & Rawlins, D.R. (1984) Template requirement for the initiation of adenovirus DNA replication. **Proc. Natl. Acad. Sci.** 81: 100-4

Challberg, M.D., Desiderio, S.V. & Kelly, T.J. (1980) Adenovirus DNA replication *in vitro*: Characterisation of a protein covalently linked to nascent DNA strands. **Proc. Natl. Acad. Sci.** 77 (9) 5105-9

Challberg, M.D., Ostrove, J.M. & Kelly, T.J. (1982) Initiation of adenovirus DNA replication: Detection of covalent complxes between nucleotide and the 80 kDa terminal protein. **J. Virol.** 41 (1) 265-70

Chan, B.S., Court, D.A., Vierula, P.J. & Bertrand, H. (1991) The kalilo linear senescence inducing plasmid of *Neurospora* is an invertron and encodes DNA and RNA polymerases. **Curr. Genet.** 20: 225-37

Chen, D-C., Yang, B-C. & Kuo, T-T. (1992) One step transformation of yeast in stationary phase. **Curr. Genet.** 21: 83-4

Chen, H., Lawrence, C.B., Bryan, S.K. & Myers, R.E. (1990) Aphidicolin inhibits DNA polymerase II of *E. coli*, an α-like DNA polymerase. **Nucleic Acid Res.** 18 (23) 7185

Chen, M. & Horowitz, M.S. (1989) Dissection of functional domains of adenovirus DNA polymerase by linker insertion mutagenesis. **Proc. Natl. Acad. Sci.** (86) 6116-20

Chen, M., Mermod, N. & Horowitz, M.S. (1990) Protein - protein interaction between adenovirus DNA polymerase and NFI mediate formation of the DNA replication preinitiation complex. **J. Biol. Chem.** 265 (10) 18634-42

Chen, X.J., Falcone, C., Bianchi, M.M. & Fukuhara, H. (1986) Sequence organisation of the circular plasmid pKD1 from the yeast *Kluyveromyces drosphilarum*. **Nucleic Acid Res.** 14 (11) 4471-9

Chou, P.Y. & Fasman, R.F. (1978) Prediction of secondary structure of proteins from their amino acid sequence. **Adv. Enzymol.** 47: 45-148

Chowdhury, C. (1991) One step mini-prep method for the isolation of plasmid DNA. **Nucleic Acid Res.** 19 (10) 2797

Clarke, L. & Carbon, J. (1978) Functional expression of cloned yeast DNA in *E. coli*: specific complementation of argininosuccinate lyase (*arg* H) mutations. **J. Mol. Biol.** 120: 517-532

Cleat, P.H. & Hay, R.T. (1989a) Co-operative interactions between NF I and adenovirus DNA binding protein at adenovirus origin of replication. **EMBO J.** 8 (6) 1841-48

Cleat, P.H. & Hay, R.T. (1989b) Kinetic analysis of NFI and its DNA binding domain with the adenovirus origin of replication. **F.E.B.S. Lett** 258 (1) 51-4

Coleman, J., Green, P.J. & Inouye, M. (1984) The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. **Cell** 37: 429-36

Cossart, P.& Sicquel-Sanzey, B. (1985) Regulation of expression of the *crp* gene of *E. coli* K12: *in vitro* study. **J. Bacteriol.** 161 (1) 454-57

Court, J.A., Griffiths, A.J.F., Kraus, S.R., Russel, P.J. & Bertrand, H. (1991) A new senescence inducing mitochondrial linear plasmid in field isolated *N.crassa* strains from India. **Curr. Genet.** 19:129-37

Crea, R., Kraszewski, A., Hirose, T. & Itakura, K. (1978) Chemical synthesis of genes for human insulin. **Proc. Natl. Acad. Sci.** 75 (12) 5765-9

Defeo-Jones, D., Huang, P.S., Jones, R.E., Haskell, K.M., Vuocolo, G.A., Hanobik, M.G., Huber, H.E. & Oliff, A. (1991) Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene product. **Nature** 352: 251-4

Delarue, M., Pock, O., Turdo, N., Meras, D. & Argos, P. (1990) An attempt to unify the structure of polymerases. **Protein Engineering** 3: 461-7

Desiderio, S.V. & Kelly, T.J. (1981) Structure of the linkage between adenovirus DNA and the 55 kDa molecular weight terminal protein. **J. Mol. Biol.** 145: 319-37

Diffley, J.F.X. & Stillman, B. (1986) Purification of a cellular ds DNA binding protein required for initiation of adenovirus DNA replication by using a rapid filter binding assay. **Mol. Cell. Biol.** 6 (5) 1363-73

DiMarchi, R., Brooke, G., Gale, C., Cracknell, V., Doel, T. & Mowat, N. (1986) Protection of cattle against foot-and-mouth disease by a synthetic peptide. **Science** 232: 639-41

Dobbs, L., Zhao, L-J.,Sripd, G. & Padmanabham, R. (1990) Mutational analysis of ss DNA template active in the *in vitro* initiation assay for adenovirus DNA replication. **Virology** 178: 43-51

Driscoll, P.C., Cyster, J.G., Campbell, I.D. & Williams, A.F. (1991) Structure of domain 1 of rat T-lymphocyte CD2 region. **Nature** 353: 762-6

Dunning, A.M., Talmud, P. & Humphries. (1988) Errors in the polymerase chain reaction. **Nucleic Acid Res.** 16 (21) 10393

Duvell, A., Hessberg-Stutzke, H., Oeser, B., Rogmann-Backwinkel, P. & Tudzynsski, P. (1988) Structural and functional analysis of mitochondrial plasmids in *Claviceps purpurea*. **Mol.Gen.Genet.** 214: 128-34

Echols, H. & Goodman, M.F. (1991) Fidelity mechanisms in DNA replication. **Ann. Rev. Biochem.** 60: 477-511

Eckert, K.A. & Kunkel, T.A. (1991) DNA polymerase fidelity and the polymerase chain reaction. **PCR Meth.Appl.** 1 (1) 17-24

Edge, M.D., Greene, A.R., Heathcliffe, G.R., Meacock, P.A., Schoch, W., Scanlon, D.B., Atkinson, T.C., Newton, C.R. & Markham, A.F. (1981) Total synthesis of a human leucocyte interferon gene. **Nature** 292: 756-62

Ellison, M.J., Kelleher, R.J. & Rich, A. (1985) Thermal regulation of β-galactosidase synthesis using antisense RNA directed against coding portion of the RNA. **J. Biol. Chem.** 260 (16) 9085-7

Emini, E.A., Hughes, J.V., Perlow, D.S. & Buger, J. (1985) Induction of hepatitis A virus neutralising antibodies by a virus-specific synthetic peptide. **J. Virol.** 55: 836-39

Enomoto, T., Lichy, J.H., Ikeda, J-E & Hurwitz, J. (1981) Adenovirus DNA replication *in vitro*: Purification of the terminal protein in a functionl form. **Proc. Natl. Acad. Sci.** 78 (11) 6779-83

Erickson, L., Beversdorf, W.D. & Pauls, K.P. (1985) Linear mitochondrial plasmid in *Brassica* has terminal protein. **Curr. Genet.** 9: 679-82

Erickson, L., Kemble, R. & Swanson, E. (1989) The *Brassica* mitochondial plasmid can be sexually transmitted. Pollen transfer of a cytoplasmic genetic element. **Mol. Gen. Genet.** 218: 419-22

Erlich, H.A. (Ed) PCR technology (1989) Stockton Press

Escarmis, C. & Salas, M. (1982) Nucleotide sequence of the early genes 3 and 4 of bacteriophage ø29. **Nucleic Acid Res.** 10 (19) 5785-90

Escarmis, C., Garcia, P., Mendez, E., Lopez, R., Salas, M. & Garcia, E. (1985) ITR and terminal proteins of the genomes of pneumococcal phages. **Gene** 30: 341-8

Escarmis, C., Guirai, D. & Salas, M. (1989) Replication of recombinant ø29 DNA molecules in *B.subtilis* protoplasts. **Virology** 169: 152-60

Euguchi, Y, Ito, T. & Tomizawa, J-I. (1991) Antisense RNA. **Ann. Rev. Biochem.** 60: 631-52

Evan, G.I., Lewis, G.K., Ramsey, G. & Bishop, J. (1985) Isolation of monoclonal antibodies specific for human *c-myc* proto-onco gene product. **Mol. Cell. Biol.** 5 (12) 3610-16

Field, J., Gronostajski, R.M. & Hurwitz, J. (1984) Properties of the adenovirus DNA polymerase. **J. Biol. Chem.** 259 (15) 9487-95

Feinberg, A. & Vogelstein, B. (1983) A technique for radiolabelling DNA restriction enzyme fragments to a high specific activity. **Anal. Biochem.** 132: 6-13

Fleer, R., Amellal, N., Yeh, P., Maury, I., Fournier, A., Bachetta, F., Badual, P., Jung, G., Hote, H.L, Becquart, J., Fukuhara, H. & Mayaux, J.F. (1991a) Stable multicopy vectors for high level secretion of recombinant human serum albumin by *Kluyveromyces* yeasts. **Biotechnology** 9: 968-78

Fleer, R., Chen, X.J., Amellal, N., Yeh, Y., Fournier, A., Guinet, F., Gault, N., Fouchier, D., Folliard, F., Fukuhara, H. & Mayaux, J.F. (1991b) High level secretion of correctly processed recombinant human IL-1β in *K. lactis*. **Gene** 107: 285-95

Fox, B.A. & Bzik, D.J. (1991) The primary structure of Plasmodium falciparum DNA polymerase δ is similar to drug sensitive δ-like viral DNA polymerases. **Mol. Biochem. Parasitol.** 49: 289-296

Francou, F. (1981) Isolation and characterisation of a linear DNA molecule in the fungus *Ascolobus immersus*. **Mol.Gen.Genet.** 184: 440-4

Fredman, J.N., Petit, S.C., Horwitz, M.S. & Engler, J.A. (1991) Linker insertion mutations in the Ad pTP that affect DNA replication activity *in vivo & in vitro*. **J. Virol.** 65 (9) 4591-7

Friefeld, B.R., Lichy, J.H., Hurwitz, J. & Horwitz, M.S. (1983) Evidence for an altered adenovirus DNA polymerase in cells infected with mutant H5ts149. **Proc. Natl. Acad. Sci.** 80: 1589-93

Friefeld, B.R., Lichy, J.H., Field, J. Guggenheimer, R.A., Krevolin, M.D., Nagata, K., Hurwitz, J. & Horwitz, M.S. (1984) The *in vitro* replication of adenovirus DNA. **Curr. Top. Micro. Immun.** 110: 220-55

Friefeld, B.R., Korn, R., De Jong, P.J., Sninsky, J.J. & Horwitz, M.S. (1985) The 140 KDa adenovirus DNA polymerase is recognised by antibodies to *E. coli* synthesized determinants predicted from an open reading frame on the adenovirus genome. **Proc. Natl. Acad. Sci.** 82: 2652-6

Fujimura, H., Hishinuma, F. & Gunge, N. (1987) Terminal segment of *K. lactis* linear DNA plasmid pGKL2 supports autonomous replication of hybrid plasmids in *S. cerevisiae*. **Curr. Genet.** 12:99-104

Fujimura, H., Yamada, T., Hishinuma, F. & Gunge, N. (1988) DNA replication *in vivo* of linear DNA killer plasmids pGKL1 and pGKL2 in *Saccharomyces cerevisiae*. **FEMS Micro.Lett.** 49: 441-4

Fukuhara, H. (1987) The RF1 gene of the killer DNA of yeast may encode a DNA polymerase. **Nucleic Acid Res.** 15 (23) 10046

Furlong, J., Meighan, M., Conner, J., Murray, J. & Clements, J.B. (1992) Methods for improved protein expression using pET vectors. **Nucleic Acid Res.** 20 (17) 4668

Garcia, E., Gomez, A., Ronda, C.,Escarmis, C. & Lopez, R. (1983) Pneumococcal bacteriophage Cp-1 contains a protein bound to the 5' termini of its DNA. **Virology** 128: 92 -104

Garcia, J.A., Pastrana, R., Prieto, I. & Salas, M. (1983) Cloning and expression in *E. coli* of the gene coding for the protein linked to the ends of *B.subtilis* phage ø29 DNA. **Gene** 21: 65-76

Garcia, J.A., Penalva, M.A., Blanco, L. & Salas, M. (1984) Template requirements for intiation of phage ø29 DNA replication *in vitro*. **Proc. Natl. Acad. Sci.** 81: 80-4

Garcia, J.A., Kolacz, K., Studnicka, G.M. & Gilmore-Hebert, M. (1988) Sequence of S1 homologous DNA in the normal maize mitochondrial genome. **Nucleic Acid Res.** 16 (9) 4169-

Garcia, P., Hermoso, J.M., Garcia, J.A., Garcia, E., Lopez, R. & Salas, M. (1986) Formation of a covalent complex between the terminal protein of Pneumococcal bacteriophage Cp-1 and 5'-dAMP. **J. Virol.** 58 (1) 31-5

Garmendia, C., Salas, M. & Hermoso, J.M. (1988) Site directed mutagenesis in the DNA linking site of bacteriophage ø29 terminal protein: isolation & characterisation of a ser^{232}-thr mutant. **Nucleic Acid Res.** 16 (13) 5727-41

Garmendia, C., Hermoso, J.M. & Salas, M. (1990) Functional domain for priming activity in phage ø29 terminal protein. **Gene** 88 (1) 73-79

Garmendia, C., Bernad, A., Esteban, J.A., Blanco, L. & Salas, M. (1992) The bacteriophage ø29 DNA polymerase, a proofreading enzyme. **J. Biol. Chem.** 267 (4) 2594-9

Garnier, J., Osguthorpe, D.J. & Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. **J. Mol. Biol.** 120: 97-120

Gassmann, M., Thommes, P., Weiser, T. & Hubscher, U. (1990) Efficient production of chicken egg yolk antibodies against a conserved mammalian protein. **FASEB. J.** 4: 2528-32

Gerendasy, D. & Ito, J. (1990) Nucleotide sequence and transcription of the right early region of bacteriophage PRD1. **J. Bacteriol.** 172 (4) 1889-1898

Giasson, L. & Lalaonde, M. (1987) Analysis of a linear plasmid isolated from the pathogenic fungus *Ceratocystis fimbriata* EII. and *Halst.* **Curr. Genet.** 11: 331-4

Gibbs, J.S., Chiou, H.C., Hall, J.D., Mount, D.W. & Retondo, M.J. (1985) Sequence and mapping analyses of the herpes simplex virus DNA polymerase gene predict a C-terminal substrate binding domain. **Proc. Natl. Acad. Sci.** 82: 7969-73

Gibbs, J.S., Weisshart, K. Digard, P., DeBruynkops, A., Knipe, D.M. & Coen, D.M. (1991) Polymerisation activity of an α-like DNA polymerase requires a conserved 3'-5' exonuclease active site. **Mol. Cell. Biol.** 11: 4786-95

Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R, Hirose, T., Kraszewski, A., Itakura, K. & Riggs, A.D. (1979) Expression in *E. coli* of a chemically synthesised gene for human insulin. **Proc. Natl. Acad. Sci.** 76: 106-10.

Gold, L. & Stormo, G.D. (1990) High level translational initiation. **Meth. Enzym.** 185: 89-93.

Gottesman, M.E., Adhya, S. & Das, A. (1980) Transcriptional antitermination by the bacteriophage λ N gene product. **J. Mol. Biol.** 140: 57-75

Gottesman, S. (1990) Minimising proteolysis in *E. coli*: Genetic solutions. **Meth. Enzym.** 185: 119-29

Gough, J.A. & Murray, N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. **J. Mol. Biol.** 166: 1-19

Gounari, F., Franesco, R.D., Schmitt, J. Vliet, P.V.D. & Stunnenberg, H. (1990) N-terminal domain of NF1 binds to DNA as a dimer and activates adenovirus DNA replication. **EMBO J.** 9 (2) 559-66

Graessmann, M. & Graessmann, A. (1989) The biological activity of early SV40 antisense RNA & DNA molecules. **Curr. Top. Micro. Immunol.** 144: 97-103

Gralla, J.D. (1990) Promoter recognition and mRNA initiation by *E. coli* RNA polymerase. **Meth. Enzym.** 185: 37-54

Grant, G. (Ed.) (1992) Synthetic Peptides. A users Guide. Freeman & Co.

Green, P.J., Pines, O. & Inouye, M. (1986) The role of antisense RNA in gene regulation. **Ann. Rev. Biochem.** 55: 569-97

Grimes, S. & Anderson, D. (1989) *In vitro* packaging of bacteriophage ø29 DNA restriction fragments and the role of the terminal protein gp3. **J. Mol. Biol.** 209: 91-100

Grunstein, M. & Hogness, D.S. (1975) Colony hybridisation: a method for the isolation of cloned DNA's that contain a specific gene. **Proc. Natl. Acad. Sci.** 72: 3961-5

Guggenheimer, R.A., Nagata, K., Kenny, M. & Hurwitz, J. (1984a) Protein primed replication of plasmids containing the terminus of the adenovirus genome. **J. Biol. Chem.** 259 (12) 7815-25

Guggenheimer, R.A., Stillman, B.W., Nagata, K., Tamanoi, F. & Hurwitz, J. (1984b) DNA sequences required for the *in vitro* replication of adenovirus DNA. **Proc. Natl. Acad. Sci.** 81: 3069-73

Guillemete, J.G., Matsushima-Hibiya, Y., Atkinson, T. & Smith, M. (1991) Expression in *E. coli* of a synthetic gene coding for horse heart myoglobin. **Protein Engineering** 4 (5) 585-92

Gunge, N. (1986) Linear DNA killer plasmids from the yeast *Kluyveromyces.* **Yeast** 2:153-62

Gunge, N. & Sakaguchi, K. (1981) Intergenic transfer of DNA killer plasmid pGKLI and pGKL2 from *K. lactis* into *S. cerevisiae* by cell fusion. **J. Bacteriol.** 147 (1) 155-60

Gunge, N. & Yamane, C. (1984) Incompatability of linear DNA killer plasmids pGKL1 and pGKL2 from *K. lactis* with mitochondrial DNA from *Saccharomyces cerevisiae.* **J. Bacteriol.** 159 (2) 533-9

Gunge, N. & Kitada, K. (1988) Replication and maintenance of the *Kluyveromyces* linear pGKL plasmids. **Eur.J.Epidemiol.** 4 (4) 409-14

Gunge, G., Tamaru, A., Ozawa, F. & Sakaguchi, K. (1981) Isolation and characterisation of linear DNA plasmids from *K. lactis* and the plasmid associated killer charcter. **J. Bacteriol.** 145 (1) 382-90

Gunge, N., Murata, K. & Sakaguchi, K. (1982) Transformation of *Saccharomyces cerevisiae* with linear DNA killer plasmids from *K. lactis.* **J. Bacteriol.** 151 (1) 462-4

Gunge, N., Murakami, K., Takesako, T. & Moriyama, H. (1990) Mating type locus dependent stability of the *Kluyveromyces* linear pGKL plasmids in *Saccharomyces cerevisiae.* **Yeast** 6:417-27

Gutierrez, C., Martin, G., Sogo, J.M. & Salas, M. (1991a) Mechanism of stimulation of DNA replication by bacteriophage ø29 ss DNA binding protein p5 **J. Biol. Chem.** 266 (4) 2104-11

Gutierrez, C., Sogo, J.M. & Salas, M. (1991b) Analysis of replicatory intermediates produced during bacteriophage ø29 DNA replication *in vitro.* **J. Mol. Biol.** 222: 983-94

Gutierrez, J., Garcia, J.A., Blanco, L. & Salas, M. (1986a) Cloning and template activity of the origins of replication of phage ø29 DNA. **Gene** 43:1-11

Gutierrez, J., Vinos, J., Prieto, I., Mendez, E., Hermoso, J.M. & Salas, M. (1986b) Signals in the ø29 DNA terminal protein template for the initiation of phage ø29 DNA replication. **Virology** 155: 474-83

Gutierrez, J., Garmendia, J. & Salas, M. (1988) Characterisation of the origin of replication of bacteriophage ø29. **Nucleic Acid Res.** 16(13) 5895-5914

Hall, J.D. (1988) Modeling functional sites in DNA polymerases. **Trends In Genet.** 4 (2) 42-46

Hallet, P., Grimshaw, A.J., Wigley, D.B. & Maxwell, A. (1990) Cloning of the DNA gyrase genes under *tac* promoter control: Overproduction of the gyrase A and B proteins. **Gene** 93: 139-42

Halling, S.M., Simons, R.W., Way, J.C., Walsh, R.B. & Kleckner, N. (1982) DNA sequence organisation of IS10-right of Tn10 and comparison with IS10 left. **Proc. Natl. Acad. Sci.** 79: 2608-12

Hames, B.D. & Rickwood, D. (1981). An introduction to gel electrophoresis of proteins. IRL Press.

Hamilton, A.J., Lycett, G.W. & Grierson, D. (1990) Antisense gene that inhibits synthesis of the hormone ethylene in transgenic plants. **Nature** 346: 284-7

Hanahan, D. (1983) Studies on transformation of *E. coli* with plasmids. **J. Mol. Biol.** 166: 557-80

Harding, N.E. & Ito, J. (1980) DNA replication of bacteriophage ø29: Characterisation of the intermediates and location of the termini of replication. **Virology** 104: 323-38

Harlow, E. & Lane, D. (1988) Antibodies: A laboratory manual. Cold Spring Harbour

Harris, M.P.G. & Hay, R.T. (1988) DNA sequences required for the initiation of adenovirus type 4 DNA replication *in vitro.* **J. Mol. Biol** 201: 57-67

Hashiba, T., Homma, Y., Hyakumachi, M. & Matsuda, I. (1984) Isolation of a DNA plasmid in the fungus *Rhizoctonia solani.* **J. Gen. Micro.** 130:2067-70

Hassin, D., Korn, R. & Horwitz, M.S. (1986) A major internal initiation site for *in vitro* translation of the adenovirus DNA polymerase. **Virology** 155: 214-24

Hay, R.T. (1985a) Origin of adenovirus DNA replication: Role of the nuclear factor I binding site *in vivo.* **J. Mol. Biol.** 186: 129-36

Hay, R.T. (1985b) The origin of adenovirus DNA replication: minimal DNA sequence requirement *in vitro*. **EMBO J.** 4 (2) 421-6

Hay, R.T. & McDougall, I.M. (1986) Viable viruses with deletions in the left inverted terminal repeats define the adenovirus origin of DNA replication. **J. Gen. Virol.** 67: 321-32

Hay, R.T., Stow, N.D. & McDougall, I.M. (1984) Replication of adenovirus mini-chromosomes. **J. Mol. Biol.** 175:493-510

Hay, R.T., Clark, L., Cleat, P.H., Harris, M.P.G., Robinson, E.C. & Watson, C.J. (1988) Requirements for initiation of adenovirus type 2 and 4 DNA replication. **Cancer Cells** 6: 71-75

Hayman, G.T. & Bolen, P.L. (1991) Linear DNA plamids of *Pichia inositovora* are associated with a novel killer toxin activity. **Curr. Genet.** 19: 129-37

Herman, R.A., Hui, H.L., Andracki, M.E., Noble, R.W., Sligar, S.G., Walder, J.A. & Walder, R.Y. (1992) Human haemoglobin expression in *E. coli*: Importance of optimal codon bias. **Biochemistry** 31 (36) 8619-28

Hermanns, J & Osiewacz, H.D. (1992) The linear mitochondrial plasmid pAL2-1 of a long lived *Podospora anserina* mutant is an invertron encoding an RNA and DNA polymerase. **Curr. Genet.** 22 (6) 491-500.

Hermoso, J.M. & Salas, M. (1980) Protein p3 is linked to the DNA of phage ø29 through a phosphester bond between serine and 5'-dAMP. **Proc. Natl. Acad. Sci.** 77 (11) 6425-8

Hermoso, J.M., Mendez, E., Soriano, F. & Salas, M. (1985) Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage ø29. **Nucleic Acid Res.** 13 (21) 7715-28

Hinnebusch, J. & Barbour, A.G. (1991) Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. **J. Bacteriol.** 173 (22) 7233-9

Hinnebusch, J., Bergstrom, S. & Barbour, A.G. (1990) Cloning and sequence analysis of linear plasmid telomeres of the bacterium *Borrelia burgdorferi*. **Molecular Microbiology** 4 (5) 811-20

Hirochika, H. & Sakaguchi, K. (1982) Analysis of linear plasmid isolated from *Streptomyces*: Association of protein with the ends of the plasmid DNA. **Plamid** 7:59-65

Hirochika, H., Nakamura, K. & Sakaguchi, K. (1984) A linear DNA plasmid from *Streptomyces rochei* with an ITR of 614bp. **EMBO J.** 3 (4) 761-6

Hirokawa, H. (1972) Transfecting DNA of the *Bacillus* bacteriophage ø29 that is protease sensitive. **Proc. Natl. Acad. Sci.** 69: 1555-9

Hishinuma, F. & Hirai, K. (1991) Genome organisation of the linear plasmid pSKL isolated from *S.kluyveri*. **Mol. Gen. Genet.** 226: 97-106

Hishinuma, F., Nakamura, K., Hirai, K., Gunge, N. & Maeda, T. (1984) Cloning and nucleotide sequences of the linear DNA killer plasmids from yeast. **Nucleic Acid Res.** 12 (9) 7581

Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. & Stuber, D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. **Biotechnology** 6 (11): 1321-5

Hoekema, A., Kastelein, R.A., Vasser, M. & De Boer, H.A. (1987) Codon replacement in the *PGK1* gene of *S. cerevisiae*: Experimental approach to study the role of biased codon usage in gene expression. **Mol. Cell. Biol.** 7 (8) 2914-24

Hoffman, C.S. & Winston, F. (1987) A 10 minute DNA preparaion from yeast efficiently releases autonomous plasmids for transformation of *E. coli*. **Gene** 57: 267-72

Honeyman, A.L. & Curvier, T.C. (1986) Isolation and characterisation of linear DNA elements from the mitochondria of *Gaeumannomyces graminis*. **Appl. Environm. Microbiol.** 52: 924-9

Hoopes, B.C. & McClure, W.R. (1985) A cII-dependent promoter is located within the Q gene of the bacteriophage λ. **Proc. Natl. Acad. Sci.** 82: 3134-8

Hopp, T.P. & Woods, K.R. (1983) Prediction of protein antigenic determinants from amino acid sequences. **Mole. Immunol.** 20: 483-9

Horowitz, M.A. (1978) Temperature sensitive DNA replication of H5ts125 adenovirus DNA *in vitro*. **Proc. Natl. Acad. Sci.** 75 (9) 4291-5

Hsieh, J-C. & Ito, J. (1988) Possible evolutionary relationship between poliovirus RNA-linked VPg and PRD1 DNA terminal protein. **Nucleic Acid Res.** 16 (9) 4180

Hsieh, J.C., Jung, G., Leavitt, M.C. & Ito, J. (1987) Primary structure of the DNA terminal protein of bacteriophage PRD1. **Nucleic Acid Res.** 15 (21) 8999-9005

Hu, C-H., Xu, F-Y., Wang, K., Pearson, A.N. & Pearson, G.D. (1992) Symmetrical adenovirus mini-chromosomes have hairpin replication intermediates. **Gene** 110: 145-50

Hui, A., Hayflick, J., Dinkelspiel, K. & de Boer, H.A (1984) Mutagenesis of the 3 bases preceding the start codon of the β-galactosidase mRNA and its effect upon translation in *E. coli*. **EMBO** J. 3: 623-9

Ikeda, J-E., Enamoto, T. & Hurwitz, J. (1981) Replication of adenovirus DNA protein complex with purified proteins. **Proc. Natl. Acad. Sci.** 78 (2) 884-8

Ikeda, J-E., Enomoto, T. & Hurwitz, J. (1982) Adenoviral protein-primed initiation of DNA chains *in vitro*. **Proc. Natl. Acad. Sci.** 79: 2442-6

Ikemura, T. (1981) Correlation between the abundance of *E. coli* transfer RNAs and the occurence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translation system. **J. Mol. Biol.** 158: 573-97

Ikemura, T. (1982) Correlation between the abundance of yeast transfer RNAs and the occurence of the respective codons in protein genes. **J. Mol. Biol.** 158: 573-97

Inouye, M. (1988) Antisense RNA: its functions and applications in gene regulation - a review. **Gene** 72: 25-34

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F. & Boyer, H.W. (1977) Expression in *E. coli* of a chemically synthesised gene for the hormone somatostatin. **Science** 198:1056-63

Ito, J. (1978) Bacteriophage ø29 terminal protein: its association with the 5' termini of the ø29 genome. **J. Virol.** 28 (3) 895-904

Ito, J. & Braithwaite, D.K. (1991) Compilation and alignement of DNA polymerase sequences. **Nucleic Acid Res.** 19 (15) 4045-57

Izant, J.G. & Weintraub, H. (1984) Inhibition of thymidine kinase gene expression by antisense RNA: A molecular approach to genetic analysis. **Cell** 36: 1007-15

Jameson, B.A. & Wolf, H. (1988) The Antigenicity Index: a novel algorithm for predicting antigenic determinants. **Comput. Appl. Biosci.** 4: 181-86

Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R.A., Nordheim, A. & Stunnenberg, H.G. (1991) Rapid and efficient purification of native histidine tagged protein expressed by recombinant vaccinia virus. **Proc. Natl. Acad. Sci.** 88: 8972-6

Jayaraman, K., Fingar, S.A., Shah, J. & Fyles, J. (1991) Polymerase chain reaction mediated gene synthesis: synthesis of a gene coding for isozyme C of horseradish peroxidase. **Proc. Natl. Acad. Sci.** 88: 4084-8

Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. & Tjian, R. (1987) A cellular DNA binding protein that activates eukaryotic transcription and DNA replication. **Cell** 48: 79-89

Joung, I., Horwitz, M.S. & Engler, J.A. (1991) Mutagenesis of conserved region I in the DNA polymerase from human adenovirus serotype 2. **Virology** 184: 235-41

Jung, G., Leavitt, M.C. & Ito, J. (1987a) Yeast killer plasmid pGKL1 encodes a DNA polymerase belonging to the family B DNA polymerases. **Nucleic Acid Res.** 15 (21) 9088

Jung, G., Leavitt, M.C., Hsieh, J.C. & Ito, J. (1987b) Bacteriophage PRD1 DNA polymerase: Evolution of DNA polymerases. **Proc. Natl. Acad. Sci.** 84: 8287-91

Kaelin, W.G., Pallas, D.C., DeCaprio, J.A., Kaye, F.J. & Livingston, D.M. (1991) Identification of cellular proteins that can interact specifically with the T/E1A binding region of the retinoblastoma gene product. **Cell** 64: 521-32

Kamer, G. & Argos, P. (1984) Primary structure comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. **Nucleic Acid Res.** 12: 7269-82

Kamper, J., Meinhardt, F., Gunge, N. & Esser, K. (1989a) *In vivo* construction of linear vectors based on killer plasmids from *K. lactis*: selection of a nuclear gene results in attachment to telomeres. **Mol. Cell. Biol.** 9 (9) 3931-7

Kamper, J., Meinhardt, F., Gunge, N. & Esser, K. (1989b) New recombinant linear DNA elements derived from *K. lactis* killer plasmids. **Nucleic Acid Res.** 17 (4) 1781

Kamper, J., Esser, K., Gunge, N. & Meinhardt, F. (1991) Heterologous gene expression on the linear DNA killer plasmid from *K. lactis*. **Curr. Genet.** 19 (2) 109-18

Kaplan, L.M., Ariga, H., Hurwitz, J. & Horwitz, M.A. (1979) Complementation of *ts* defect in H5*ts*125 adenovirus DNA replication *in vitro*. **Proc. Natl. Acad. Sci.** 76 (11) 5534-8

Katayose, Y., Kajiwara, S. & Shishido, K. (1990) The basidiomycete *Lentinus edodes* linear mitochondrial DNA plasmid contains a segment exhibiting a high ARS activity in *S. cerevisiae*. **Nucleic Acid Res.** 18 (6) 1395-1400

Katayose, Y., Yui, Y., Kajiwara, S. & Shishido, K. (1992) A wide distribution of linear DNA plasmids in the basidiomycetes *L. edodes* and *P. ostreatis*. **J. Gen. Appl. Micro.** 38 (3) 289-92

Keen, C.L., Mendelovitz, S. Cohen, G. Aharonowitz, Y. & Roy, K.L. (1988) Isolation and characterisation of a linear DNA plasmid from *Streptomyces clavuligerus*. **Mol. Gen. Genet.** 212:172-6

Kelly, T.J., Wold, M.S. & Li, J. (1988) Initiation of viral DNA replication. **Adv. Vir. Res.** 34: 1-42

Kemble, R.J. & Thompson, R.D. (1982) S1 and S2, the linear mitochondrial DNAs present in a male sterile line of maize, posses terminally attached proteins. **Nucleic Acid Res.** 10 (24) 8181-90

Kempken, F., Meinhardt, F. & Esser, K. (1989) *In organello* replication and viral affinity of linear extrachromosomal DNA of the ascomycete *Ascobolus immersus*. **Mol. Gen. Genet.** 218: 523-30

Kenny, M.K. & Hurwitz, J. (1988) Initiation of adenovirus DNA replication. **J. Biol. Chem.** 263 (20) 9809-17

Kenny, M.K., Balogh, L.A. & Hurwitz, J. (1987) Initiation of adenovirus DNA replication. **J. Biol. Chem.** 263 (20) 9801-17

Kiefhaber, T., Rudolph, R., Kohler, K.K. & Buchner, J. (1991) Protein aggregation *in vitro* and *in vivo*: A quatitative model of the kinetic competition between folding and aggregation. **Biotechnology** 9: 825-30

Kikuchi, Y., Hirai, K. & Hishinuma, F. (1984) The yeast linear DNA killer plasmids, pGKL1 and pGKL2, posses terminally attached proteins. **Nucleic Acid Res.** 12 (14) 5685-92

Kikuchi, Y., Hirai, K., Gunge, N. & Hishinuma, F. (1985) Hairpin plasmid - a novel linear DNA of perfect hairpin structure. **EMBO J.** 4 (7) 1881-6

Kim, S.K. & Wold, B.J. (1985) Stable reduction of thymidine kinase activity in cells expressing high levels of antisense RNA. **Cell** 42:129-38

Kim, W.K., Whitmore, E. & Klassen, G.R. (1990) Homologous linear plasmids in mitochondria of 3 species of wheat bunt fungi, *Tilletia caries*, *T. laevis* and *T. controversa*. **Curr. Genet.** 17: 229-33

Kinashi, H. & Shinagi, M. (1987) Detection of giant linear plasmids in antibiotic producing strains of *Streptomyces* by the OFAGE technique. **J. Antibiot.** 40 (6) 913-6

Kinashi, H., Shimaji, M. & Sakai, A. (1987) Giant linear plasmids in Streptomyces which code for antibiotic biosynthesis genes. **Nature** 328:454-6

Kistler, H.C. & Leong, S.A. (1986) Linear plasmid like DNA in the plant pathogenic fungus *Fusarium oxysporum f.sp. conglutinans*. **J. Bacteriol.** 167 (2) 587-93

Kitada, K. & Gunge, N (1988) Palindrome-hairpin linear plasmids possessing only a part of the ORF1 gene of the yeast killer plasmid pGKL1. **Mol. Gen. Genet.** 215:46-52

Kitada, K. & Hishinuma, F. (1987) A new linear DNA plasmid isolated from the yeast *Saccharomyces kluyveri*. **Mol. Gen. Genet.** 206: 377-81

Knopf, C.W. (1987) The HSV type 1 DNA polymerase gene: site of phosphonoacetic acid resistance mutation in strain angelotti is highly conserved. **J. Gen. Virol.** 68: 1429-33

Kobayashi, H., Matsumoto, K., Misawa, S., Miura, K-I & Hirokawa, H. (1989) An inhibitory effect of RGD peptide on protein-priming reaction of bacteriophages ø29 and M2. **Mol. Gen. Genet.** 220: 8-11

Kobayashi, H., Kitabayashi, K., Matsumoto, K. & Hirokawa, H. (1991) Primer protein of bacteriophage M2 exposes the RGD receptor site upon linking the first dNTP. **Mol. Gen. Genet.** 226: 65-69

Kohno, T., Carmichael, D.F., Sommer, A. & Thompson, R.C. (1990) Refolding of recombinant proteins. **Meth. Enzym.** 185: 187-210

Koprowski, H. & Melchers, F. (Ed.) (1986) Peptides as Immunogens. **Curr. Top. Micro. Immunol.** Volume 130 Springer-Verlag.

Kornberg, A. & Baker, T. (1992) DNA replication. 2nd Ed. Freeman & Co.

Kotula, L. & Curtis, P.J. (1991) Evaluation of foreign gene codon optimisation in yeast: expression of a mouse Ig Kappa chain. **Biotechnology** 9: 1386-90

van der Krol, A.R., Mol, J.M. & Stuije, A. (1988) Antisense genes in plants: An overview. **Gene** 72: 45-54

Kuzmin, E.V. & Levchenko, I.V. (1987) S1 plasmid from cms-S-maize mitochondria encodes a viral type DNA polymerase. **Nucleic Acid Res.** 15 (16): 6758

Kuzmin, E.V., Levchenko, I.V. & Zaitseva, G.N. (1988) S2 plasmid from cms-S-maize mitochondria potentially encodes a specific RNA polymerase. **Nucleic Acid Res.** 16 (9) 4177

Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying the hydropathic charcter of a protein. **J. Mol. Biol.** 157: 105-32

Larder, B.A., Kemp, S.D. & Darby, G. (1987) Related functional domains in virus DNA polymerases. **EMBO J** 6 (1) 169-75

Le Grice, S.F.J. & Gruninger-Leitch, F. (1990) Rapid purification of homodimer and heterodimer HIV-1 reverse transcriptase by metal chelate affinity chromatography. **Eur. J. Biochem.**: 307-14

Leavitt, M.C. & Ito, J. (1987) Nucleotide sequence of *Bacillus* phage Nf terminal protein gene. **Nucleic Acid Res.** 15 (13) 5251-60

Leavitt, M.C. & Ito, J. (1989) T5 DNA polymerase: structural-functional relationships to other DNA polymerases. **Proc. Natl. Acad. Sci.** 86: 4465-9

Lechner, R.L. & Kelly, T.J. (1977) The structure of replicating Adenovirus 2 DNA molecules. **Cell** 12: 1007-20

Leegwater, P.A.J., van Driel, W. & van der Vliet, P.C. (1985) Recognition site of NF I, a sequence specific DBP from HeLa cells that stimulates adenovirus DNA replication. **EMBO J** 4: 1515-21

Leith, I.R., Hay. R.T. & Russel, W.C. (1989) Adenovirus sub-viral particles and cores can support limited DNA replication. **J. Gen. Virology.** 70:3235-48

Leon, P., Walbot, V. & Bedinger, P. (1989) Molecular analysis of the linear 2.3 kb plasmid of maize mitochondria: apparent capture of tRNA genes. **Nucleic Acid Res.** 17 (11) 4089-99

Lerner, R.A. (1982) Tapping the immunological repetoire to produce antibodies of predetermined specificty. **Nature** 299: 592-6

Levings, C.S. & Sederoff, R.R. (1983) Nucleotide sequence of the S2 mitochondrial DNA from the S-cytoplasm of maize. **Proc. Natl. Acad. Sci.** 50: 4055-9

Lichy, J.H., Horowitz, M.S. & Hurwitz, J. (1981) Formation of a covalent complex between the 80 kDa adenovirus terminal protein and 5'-dCMP *in vitro*. **Proc. Natl. Acad. Sci.** 78 (5) 2678-82

Lichy, J.H., Field, J., Horwitz, M.A. & Hurwitz, J. (1982) Separation of adenovirus terminal protein precursor from its associated DNA polymerase: Role of both proteins in the initiation of adenovirus DNA replication. **Proc. Natl. Acad. Sci.** 79: 5225-9

Ligon, J.M., Bolen, P.L., Hill, D.S., Bothast, R.J. & Kurtzman, C.P. (1989) Physical and biological characterisation of linear DNA plasmids of the yeast *Pichia inositovora* **Plasmid** 21: 185-94

Lindenbaum, J.O., Field, J. & Hurwitz, J. (1986) The adenovirus DNA binding protein and adenovirus DNA polymerase interact to catalyse elongation of primed DNA templates. **J. Biol. Chem.** 261 (22) 10218-27

Lippe, R. & Graham, F.L. (1989) Adenoviruses with non-identical terminal sequences are viable. **J. Virol.** 63 (12) 5133-41

Ljungquist, I., Breitholtz, A., Brink-Nilsson, H., Moks, T., Uhlen, M. & Nilsson, B. (1989) Immobilization and affinity purification of recombinant proteins using histidine peptide fusions. **Biochemistry** 186: 563-9

Longley, M.J., Bennet, S.E. & Mosbaugh, D.W. (1990) Characterisation of the 5' to 3' exonuclease associated with *Thermus aquaticus* DNA polymerase. **Nucleic Acid Res.** 18 (24) 7317-20

Lopez, R., Ronda, C., Garcia, P., Escarmis, C. & Garcia, E. (1984) Restriction cleavage maps of the DNAs of *Streptococcus pneumoniae* bacteriophages containing protein covalently bound to their 5' ends. **Mol. Gen. Genet.** 197: 67-74

Louvencourt, L., Fukuhara, H., Heslot, H. & Wesolowski, M. (1983) Transformation of *K. lactis* by killer plasmid DNA. **J. Bacteriol.** 154 (2) 737-42

Ma, C. & Simons, R.W. (1990) The IS10 antisense RNA blocks ribosome binding at the transposase intiation site. **EMBO J.** 9 (4) 1267-74

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbour Laboratory, USA.

Marcy, A.I., Hwang, C.B.C., Ruffner, K.L. & Coen, D.M. (1990) Engineered HSV DNA polymerase point mutations: the most highly conserved region shared amoung α-like DNA polymerases is involved in substrate recognition. **J.Virol.** 64 (12) 5883-90

Marsh, J.L., Erfle, M. & Wykes, E.J. (1984) The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. **Gene** 32: 481-5

Martin, G. & Salas, M. (1988) Characterisation and cloning of gene 5 of *B.subtilis* phage ø29. **Gene** 67: 193-201

Martin, G., Lazaro, J.M., Mendez, E. & Salas, M. (1989) Characterisation of phage ø29 protein p5 as single stranded DNA binding protein: Function in ø29 DNA-protein p3 replication. **Nucleic Acid Res.** 17 (10) 3663-72

Matsumoto, K., Saito, T. & Hirokawa, H. (1983) *In vitro* initiation of bacteriophage ø29 and M2 DNA replication: Genes required for formation of a complex between the terminal protein and 5'dAMP. **Mol. Gen. Genet.** 191: 26-30

Matsumoto, K., Saito, T., Kim, C.I., Ando, T. & Hirokawa, H. (1984) Bacteriophage ø29 DNA replication *in vitro*: participation of the terminal protein and the gene 2 product in elongation. **Mol. Gen. Genet.** 196: 381-6

Matsumoto, K., Kim, C.L., Urano, S., Ohashi, M. & Hirokawa, H. (1986) Aphidicolin resistant mutants of bacteriophage ø29: genetic evidence for altered DNA polymerase. **Virology** 152: 32-8

Matsumoto, K., Takano, H., Kim, C.I. & Hirokawa, H. (1989) Primary structure of bacteriophage M2 DNA polymerase: conserved segments within protein-priming DNA polymerases and DNA polymerase I of *E. coli.* **Gene** 84: 247-55

Matsumoto, K., Kim, C.I., Kobayashi, H., Kanehiro, H. & Hirokawa, H. (1990) Aphidicolin-resistant DNA polymerase of bacteriophage ø29: APH71 mutant is hypersensitive to phosponoacetic acid and buPdGTP. **Virology** 178:337-9

McGarry, T.J. & Lindquist, S. (1986) Inhibiton of heat shock protein synthesis by heat inducible antisense RNA. **Proc. Natl. Acad. Sci.** 83: 399-403

McNeel, D.G. & Tamanoi, F. (1991) Terminal region recognition factor I, a DNA binding protein recognises the ITR's of the pGKL linear plasmids. **Proc. Natl. Acad. Sci.** 88: 11398-402

Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A. & Smith, L.M. (1991) A universal method for the direct cloning of PCR amplified nucleic acid. **Biotechnology** 9: 657-63

Meinhardt, F. & Esser, K. (1984) Linear extrachromosomal DNA in the morel *Morchella conica.* **Curr. Genet.** 8: 15-18

Meinhardt, F. & Esser, K. (1987) The plasmids of the morels: characterisation and prerequisites for vector development. **Appl. Micro. Biotechn.** 27: 276-282

Meinhardt, F., Kempken, F. & Esser, K. (1986) Proteins are attached to the ends of a linear plasmid in the filamentous fungus *Ascobolus immersus*. **Curr. Genet.** 11:243-6

Meinhardt, F., Kempken, F., Kamper, J. & Esser, K. (1990) Linear plasmids amoung eukaryotes: Fundementals and applications. **Curr. Genet.** 17: 89-95

Mellado, R.P., Penalva, M.A., Inciarte, M.R. & Salas, M. (1980) The protein covalently linked to the 5' termini of the DNA of *B.subtilis* phage ø29 is involved in the initiation of DNA replication. **Virology** 104: 84-96

Melton, D.A. (1985) Injected antisense RNAs specifically block mRNA translation *in vivo*. **Proc. Natl. Acad. Sci.** 82: 144-8

Miyashita, S., Hirochika, H., Ikeda, J. & Hashiba, T. (1990) Linear plasmid DNAs of the plant pathogenic fungus *Rhizoctonia solani* with unique terminal structures. **Mol. Gen. Genet.** 220:165-71

Mogen, K.L., Siegel, M.R. & schardl, C.L. (1991) Linear DNA plasmids of the perenial rye grass choke pathogen, *Epichloe typhina* (*Clavicepitaceae*). **Curr. Genet.** 20: 519-26

Morrison, A., Bell, J.B., Kunkel, T.A. & Sugino, A. (1991) Eukaryotic DNA polymerase amino acid sequence required for 3' to 5' exonuclease activity. **Proc. Natl. Acad. Sci.** 88: 9473-7

Mount, R.C., Jordan, B.E. & Hadfield, C. (1992) Transformation of lithium treated yeast cells and the selection of auxotrophic and dominant markers. Submited for publication, in Methods In Yeast Genetics And Molecular Biology. Evans, I. (Ed) Humana Press

Mul, Y.M. & van der Vliet, P.C. (1992) NF I enhances adenovirus DNA replication by increasing the stability of a pre-initiation complex. **EMBO J.** 11 (2) 751-60

Myers, C.J Griffiths, A.J.F. & Bertrand, H. (1989) Linear kalilo DNA is a *Neurospora* mitochondial plasmid that integrates into the mitochondrail DNA. **Mol. Gen. Genet.** 220: 113-20

Nagata, K., Guggenheimer, R.A. Enomoto, T., Lichy, J.H. & Hurwitz, J. (1982) Adenovirus DNA replication *in vitro*: Identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. **Proc. Natl. Acad. Sci.** 79: 6438-42

Nagata, K., Guggenheimer, R.A. & Hurwitz, J. (1983) Specific binding of a cellular DNA replication protein to the origin of replication of adenovirus DNA. **Proc. Natl. Acad. Sci.** 80: 6177-81

Nakano, R., Zhao, L.J. & Padmanabhan, R. (1991) Overproduction of adenovirus DNA polymerase and preterminal protein in HeLa cells. **Gene** 105: 173-8

Nambiar, K.P., Stackhouse, J., Stauffer, D.M., Kennedy, W.P., Eldredge, J.K. & Benner, S.A. (1984) Total synthesis and cloning of a gene coding for the ribonuclease S protein. **Science** 223: 1299-1301

Nambiar, P., Stackhouse, J., Presnell, S.R. & Benner, S.A. (1987) Expression of bovine pancreatic ribonuclease A in *E. coli*. **Biochemistry** 163: 67-71

Nilsson, B. & Abrahmsen, L. (1990) Fusions to *Staphylococcal* protein A. **Meth. Enzym.** 185: 144-61

Nilsson, B., Abrahmsen, L. & Uhlen, M. (1985) Immobilisation and purification of enzymes with *Staphylococcal* protein A gene fusion vectors. **EMBO** J 4: 1075-80

O'Neill, E.A. & Kelly, T.J. (1988) Purification and characterisation of NF 3 (Origin recognition protein C) a sequence specific DNA binding protein required for efficient initiation of adenovirus DNA replication. **J. Biol. Chem.** 263 (2) 931-7

O'Neill, E.A., Fletcher, C., Burrow, C.R., Heintz, N., Roeder, R.G. & Kelly, T.J. (1988) Transcription factor OTF-1 is functionally identical to the DNA replication factor NF-III. **Science** 1210-1214

Oeser, B. (1988) S2 plasmid from *Zea mays* encodes a specific RNA polymerase: an alternative alignement. **Nucleic Acid Res.** 16 (17) 8729

Oeser, B. & Tudzynski, P. (1989) The linear mitochondrial plasmid pCIK1 of the phytopathogenic fungus *Claviceps purpurea* may code for a DNA polymerase and an RNA polymerase. **Mol. Gen. Genet.** 217:132-40

Okamoto, K. & Freundlich, M. (1986) Mechanism for the autogenous control of the *crp* operon: Transcriptional inhibition by a divergent RNA transcript. **Proc. Natl. Acad. Sci.** 83: 5000-4

Ostrove, J.M., Rosenfeld, P., Williams, J. & Kelly, T.J. (1983) *In vitro* complementation as an assay for purification of adenovirus DNA replication proteins. **Proc. Natl. Acad. Sci.** 80: 935-9

Otero, M. & Salas, M. (1989) Regions at the C-terminal end of bacteriophage ø29 protein p6 required for DNA binding and activity in ø29 DNA replication. **Nucleic Acid Res.** 17 (12) 4567-77

Paces, V., Vlcek, C., Urbanek, P. & Hostomsky, Z. (1985) Nucleotide sequence of the major early region of *B.subtilis* phage PZA, a close relative of ø29. **Gene** 38: 45-56

Paillard, M., Sederoff, R.R. & Levings, C.S. (1985) Nucleotide sequence of the S1 mitochondrial DNA from the S cytoplasm of maize. **EMBO J.** 4(5) 1125-8

Palmer, J.D., Shields, C.R., Cohen, D.B. & Orton, T.J. (1983) An unusual mitochondrial DNA plasmid in the geneus *Brassica*. **Nature** 301:725-8

Panchal, C.J., Meacher, C., Oostrom, J.V. & Stewart, G.G. (1985) Phenotypic expression of *K. lactis* killer toxin against *Saccharomyces spp*. **Appl. Env. Micro.** 50 (2) 257-260

Patarroyo, M.E., Romero, P., Torres, M.L., Clavijo, P., Moreno, A., Martinez, A., Rodriguez, R., Guzman, F. & Cabezas, E. (1987) Induction of protective immunity against experimental infection with maleria using synthetic peptides. **Nature** 328: 629-32

Penalva, M.A. & Salas, M. (1982) Initiation of phage ø29 DNA replication *in vitro*: Formation of a covalent complex between the terminal protein, p3, and 5'-dAMP. **Proc. Natl. Acad. Sci.** 79: 5522-6

Pestka, S., Dougherty, B.L., Jung, V., Hotta, K. & Pestka, R.K. (1984) Anti mRNA: specific inhibition of translation of single mRNA molecules. **Proc. Natl. Acad. Sci.** 81: 7525-8

Pettit, S.C., Horwitz, M.S. & Engler, J.A. (1988) Adenovirus precurser terminal protein synthesized in COS cells from cloned DNA is active in DNA replication *in vitro*. **J. Virol.** 62 (2) 496-500

Pettit, S.C., Horwitz, M.S. & Engler, J.A. (1989) Mutations of the precursor to the terminal protein of adenovirus serotypes 2 and 5. **J. Virol.** 63 (12) 5244-50

Pisani, F., Martino, C.D. & Rossi, M. (1992) A DNA polymerase from the archeon *Sulfolobus solfataricus* shows sequence similarity to family B DNA polymerases. **Nucleic Acid Res.** 20 (11) 2711-16

Pizzagalli, A., Valsanini, P. & Lucchini, G. (1988) DNA polymerase I gene of *S. cerevisiae*: nucleotide sequence, mapping of a temperature sensitive mutation and protein homology with other DNA polymerases. **Proc. Natl. Acad. Sci.** 85: 3772-6

Plasterk, R.H.A., Simon, M.I. & Barbour, A.G. (1985) Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii*. **Nature**:318: 257

Posnett, D.N., McGrath, H. & Tam, J.D. (1988) A novel method for producing anti-peptide antibodies. **J. Biol. Chem.** 263 (4) 1719-25

Prieto, I., Lazaro, J.M., Garcia, J.A., Hermoso, J.M. & Salas, M. (1984) Purification in a functional form of the terminal protein of *Bacillus subtilis* phage ø29. **Proc. Natl. Acad. Sci.** 81: 1639-43

Pring, R., Levings, C.S. & Timothy, D.H. (1977) Unique DNA associated with mitochondria in the S-type cytoplasm of male sterile maize. **Proc. Natl. Acad. Sci.** 74 (7) 2904-8

Pring, D.R., Conde, M.F., Schertz, K.F. & Levings, C.S. (1982) Plasmid-like DNAs associated with mitochondria of cytoplasmic male sterile Sorghum. **Mol. Gen. Genet.** 186: 180-4

Pruijn, G.J.M., van Driel, W. & van der Vliet, P.C. (1986) NF3, a novel sequence-specific DNA binding protein from HeLa cells stimulating adenovirus DNA replication. **Nature** 322: 656-9

Pruijn, G.J.M., van der Driel, W., Miltenburg, R.T.V. & van der Vliet, P.C. (1987) Promotor and enhancer elements containing a conserved sequence motif are recognised by NFIII - a protein stimulating adenovirus DNA replication. **EMBO J.** 6 (12) 5771-8

Pruijn, G.J.M., van der Vliet, P.C., Dathan, N.A. & Mattay, I.W. (1989) Anti-OTF 1-antibodies inhibit NFIII stimulation of *in vitro* adenovirus DNA replication. **Nucleic Acid Res.** 17 (5) 1845-63

Qin, H., Welker, D.L. & Youssef, N.N. (1993) Isolation and characterisation of a linear plasmid from the entomopathogenic fungus *Ascosphaera apis*. **Plasmid** 29 (1) 19-30

Rawlins, D.R., Rosenfeld, P.J., Wides, R.J., Challberg, M.D & Kelly, T.J. (1984) Structure and function of adenovirus origin of replication. **Cell** 37: 309-19

Reddy, M.K., Weitzel, S.E. & Hippel, P.H.V. (1992) Processive proofreading in intrinsic to T4 DNA polymerase. **J. Biol. Chem.** 267 (20) 14157-66

van Regenmortal, M.H., Briand, J.P., Muller, S. & Plane, S. (1989) **Synthetic peptides as antigens.** Elsevier, Oxford.

Reha-Krantz, L.J. (1992) Are there highly conserved DNA polymerase 3'-5' exonuclease motifs. **Gene 112:** 133-7

Rekosh, D.M.K., Russell, W.C. & Bellet, A.J.D. (1977) Identification of a protein linked to the ends of adenovirus DNA. **Cell** 11: 283-95

Rekosh, D., Lindenbaum, J., Brewster, J., Mertz, L.M., Hurwitz, J. & Prestine, L. (1985) Expression in *E. coli* of a fusion protein product containing a region of the adenovirus DNA polymerase. **Proc. Natl. Acad. Sci.** 82: 2354-8

Remenick, J., Kenny, M.K. & McGowan, J.J. (1988) Inhibition of adenovirus DNA replication by vesicular stomatitis virus leader RNA. **J. Virol.** 62 (4) 1286-92

Rhodes, A. & James, W. (1990) Inhibition of HIV replication in cell cultures by endogenously added antisense RNA. **J. Gen. Virol.** 71: 1965-74

Rhodes, A. & James, W. (1991) Inhibition of heterologous strains of HIV by antisense RNA. **AIDS** 5: 145-51

Rijnders, A.W.M., Bergan, B.G.M.V., Vliet, P.C.V.D. & Sussenbach, J.S. (1983) Specific binding of the adenovirus terminal protein precursor-DNA polymerase complex to the origin of DNA replication. **Nucleic Acid Res.** 11 (24) 8777-83

Rittner, K. & Sczakiel, G. (1991) Identification and analysis of antisense RNA target regions of the human immunodeficiency virus type 1. **Nucleic Acid Res.** 19 (7) 1421

Robert, L.S., Donaldson, P.A., Ladaique, C., Altosaar, I., Arnison, P.G. & Fabijanski, S.F. (1989) Antisense RNA inhibition of β-glucuronidase gene expression in transgenic tobacco plants. **Plant Mol. Biol** 13: 399-409

Robinson, A.J. Younghusband, H.B. & Bellet, A.J.D. (1973) A circular DNA-protein complex from adenoviruses. **Virology** 56: 54-69

Robinson, M.M. Royer, J.C. & Horgen, P.A. (1991) Homology between mitochondrial DNA of *Agaricus bisporus* and an internal portion of a linear mitochondrial plasmid. **Curr. Genet.** 19: 495-502

Rohe, M., Schrage, K. & Meinhardt, F. (1991) The linear plasmid pMC3-2 from *Morchella conica* is structurally related to adenovirus. **Curr. Genet.** 20:527-33

Romanos, M.A., Scorer, C.A. & Clare, J.J. (1992) Foreign gene expression in yeast: a review. **Yeast** 8: 423-88

Romanos, M.A. & Boyd, A. (1988) A transcriptional barrier to expression of cloned genes of the linear plasmid K1 of *Kluyveromyces lacis*: evidence that native K1 has novel promotors. **Nucleic Acid Res.** 16 (15) 7333-50

Romero, A., Lopez, R., Lurz, R. & Garcia, P. (1990) Temperate bacteriophages of *S. pneumoniae* that contain protein covalently linked to the 5' ends of their DNA. **J. Virol.** 64 (10) 5149-55

Rosenberg, A.H., Lade, B.N., Chui, D.S., Lin, S.W., Dunn, J.J. & Studier, F.W. (1987) Vectors for selective expression of cloned DNA's by T7 RNA polymerase. **Gene** 56: 125-35

Rosenberg, U.B., Preiss, A., Seifert, E., Hackle, H. & Knipple, D.C. (1985) Production of phenocopies by *kruppel* antisense RNA injection into *Drosophila* embryos. **Nature** 313: 702-6

Rosenfeld, P.J. & Kelly, T.J. (1986) Purification of NF1 by DNA recognition site affinity chromatography. **J. Biol. Chem.** 261 (3) 1398-1408

Rosenfeld, P.J., O'Neill, E.A., Wider, R.J. & Kelly, T.J. (1987) Sequence specific interactions between cellular DNA binding proteins and adenovirus origin of DNA replication. **Mol. Cell. Biol.** 7: 875-86

Ruscitti, T., Polayse, D.A., Karu, A.E. & Linn, S. (1992) Selective immunoneutralisation of the multiple activities of *E. coli* DNA polymerase I supports the model for separate active sites and indicates a complex 5'-3' exonuclease. **J. Biol. Chem.** 267 (24) 16806-11.

Sakaguchi, K. (1990) Invertrons, a class of structurally and functionally related genetic elements that include linear DNA plasmids, transposable elements and genomes of Adenovirus-type viruses. **Micro. Revs.** 54 (1) 66-74

Salas, M. (1983) A new mechanism for the initiation of replication of ø29 and Adenovirus DNA: Priming by the terminal proteins. **Curr. Top. Micro. Immun.** 109: 89-106

Salas, M. (1988) Initiation of DNA replication by primed proteins: Bacteriophage ø29 and its reletives. **Curr. Top. Micro. Immun.** 136: 71-88

Salas, M. (1991) Protein priming of DNA replication. **Ann. Rev. Biochem.** 60: 39-71

Salas, M., Mellado, R.P. & Vinuela, E. (1978) Characterisation of a protein covalently linked to the 5' termini of the DNA of *Bacillus subtillis* phage ø29. **J. Mol. Biol.** 119:269-91

Samac, D.A. & Leong, S.A. (1988) Two linear plasmids in mitochondria of *Fusarium solani f.sp.cucurbitae.* **Plasmid** 19: 57-67

Samac, D.A. & Leong, S.A. (1989) Characterisation of the termini of the linear plasmids from *Nectria Haematococca* and their use in construction of an autonomously replicating transformation vector. **Curr. Genet.** 16:187-94

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain termination inhibitors. **Proc. Natl. Acad. Sci.** 74: 5463-7

Sassenfeld, H. M. & Brewer, S.J. (1984) A peptide fusion designed for the purification of recombinant proteins. **Biotechnology** 2: 76-81

Saumitou-Laprade, P., Pannenbecker, G., Jean, R. & Michaelis, G. (1989) A linear 10.4 kb plasmid in the mitochondria of *Beta maritima.* **Curr. Genet.** 16: 181-6

Savilahti, H. & Bamford, D.H. (1986) Linear DNA replication: ITR of 5 closely related *E. coli* bacteriophages. **Gene** 49: 199-205

Savilahti, H. & Bamford, D.H. (1987) The complete nucleotide sequence of the left very early region of *E. coli* bacteriophage PRD1 coding for the terminal protein and the DNA polymerase. **Gene** 57: 121-30

Savilahti, H., Caldentey, J. & Bamford, D.H. (1989) Bacteriophage PRD1 terminal protein: Expression of gene VIII in *E. coli* and purification of the functional P8 product. **Gene** 85 (1) 45-52

Savilahti, H., Caldentey, J. Lundstrom, K., Syvaoja, J.E. & Bamford, D.H. (1991) Over-expression, purification and characterisation of *E. coli* bacteriophage PRD1 DNA polymerase. **J. Biol. Chem.** 266 (28) 18737-44

Schaack, J., Schedl, P. & Shenk, T. (1990a) Topoisomerase I and II cleavage of adenovirus DNA *in vivo*: both topoisomerase activities appear to be required for adenovirus DNA replication. **J. Virol.** 64 (1) 78-85

Schaak, J., Schedl, P. & Shank, T. (1990b) Transcription of Adenovirus and HeLa cells genes in the presence of drugs that inhibit topoisomerase I and II function. **Nucleic Acid Res.** 18 (6) 1499

Schaffrath, R., Stark, M.J.R., Gunge, N. & Meinhardt, F. (1992) *Kluyveromyces lactis* killer system: ORF1 of pGKL2 has no function in immunity expression and is dispensible for killer plasmid replication and maintenance. **Curr. Genet.** 21: 357-63

Scharltd, C.L., Pring, D.R. & Lonsdale, D.M. (1985) Mitochondrial DNA rearrangements associated with fertile revertants of S-type male sterile maize. **Cell** 43: 361-8

Schechter, N.M., Davies, W. & Anderson, C.W. (1980) Adenovirus coded DNA BP. Isolation, physical properties and effects of proteolytic digestion. **Biochemistry** 19: 2802-19

Schein, K.H. (1989) Production of soluble recombinant proteins in bacteria. **Biotechnology** 7: 1141-7

Schena, M., Picard, D. & Yamamoto, K.R. (1991) Vectors for constitutive and inducible gene expression in yeast. **Meth. Enzymol.** 194:389-398.

Schmitt, E., Brown, T.A., Trumpower, B.L. (1990) A rapid and simple method for the preparation of RNA from *S. cerevisiae.* **Nucleic Acid Res.** 18 (10) 3091

Schoner, B.E., Belagaje, R.M. & Schoner, R.G. (1990) Enhanced translational efficiency with two cistron expression system. **Meth. Enzymol.** 185: 94-103

Serrano, M., Gutierrez, J., Prieto, I., Hermoso, J.M. & Salas, M. (1989) Signals at the ø29 DNA replication origin for protein p6 binding and activity. **EMBO J** 8 (6) 1879-85

Sharp, P.A., Moore, C. & Haverty, J.L. (1976) The infectivity of adenovirus 5 DNA-protein complex. **Virology** 75:442-56

Sharp, P.M. & Li, W.H. (1986) Codon usage in regulatory genes in *E. coli* does not reflect selection for rare codons. **Nucleic Acid Res.** 14 (19) 7737-49

Sharp, P.M. & Li, W-H. (1987) The codon adaption index - a measure of directional synonymous codon usage bias and its potential applications. **Nucleic Acid Res.** 15 (3) 1281-95

Shepherd, H.S. (1992) Linear non-mitochondrial plasmids of *Alternaria alternata*. **Curr. Genet.** 21:169-72

Shepherd, H.S., Ligon, J.M., Kurtzman, C.P. & Bolen, P.C. (1987) Cryptic DNA plasmids of the heterothallic yeast *S. crataegenesis*. **Curr. Genet.** 12: 297-304

Shih, M.F., Watabe, K. & Ito, J. (1982) *In vitro* complex formation between bacteriophage ø29 terminal protein and deoxyribonucleotide. **Biochem. Biophys. Res. Comm.** 105 (3) 1031-6

Shih, M.F., Watabe, K., Yoshikawa, H. & Ito, J. (1984) Antibodies specific for the ø29 terminal protein inhibit the initiation of DNA replication *in vitro*. **Virology**. 133: 56-64

Shinagawa, M. & Padmanabhan, R. (1979) Nucleotide sequence at the ITR of adenovirus type 2. **Biochem. Biophys. Res. Comm.** 87 (3) 671-8

Shine, J., Fettes, I., Lan, N.C.Y., Roberts, J. & Baxter, J.D. (1980) Expression of cloned β-endorpin gene sequence by *E. coli*. **Nature** 285: 456-61.

Shu, L., Horowitz, M.S. & Engler, J.A. (1987) Expression of enzymically active adenovirus DNA polymerase from cloned DNA requires sequences upstream of the main ORF. **Virology** 161: 520-6

Simons, R.W. & Kleckner, N. (1983) Translational control of IS10 transposition. **Cell** 34: 683-91

Simons, R.W. (1988) Naturally occuring antisense RNA control-a brief review. **Gene** 72: 35-44

Simons, R.W., Hoopes, B.C., McClure, W.R. & Kleckner, N. (1983) Three promoters near the termini of IS10: pIN, pOUT and pIII. **Cell** 34: 673-82

Smith, C.J.S., Watson, C.F., Ray, J., Bird, C.R., Morris, P.C., Schuch, W. & Grierson, D. (1988) Antisense RNA inhibition of polygalacturonidase gene expression in transgenic tomatoes. **Nature** 334: 724-6

Smith, D.B. & Corcoran, L.M. (1990) Expression and purification of GST fusion proteins. **Curr. Prot. Mol. Biol.** Unit 16.7 Supplement 10 Ed.Ausubel, F.M., Brent, R., Kingston, Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. Greene Publishing/Wiley Interscience.

Smith, D.B. & Johnson, K.S. (1988) Single step purification of polypeptides in *E. coli* as fusions with GST. **Gene** 67: 31-40

Soengas, M.S., Esteban, J.A., Lazaro, J.M., Bernad, A., Blasco, M.A., Salas, M. & Blanco, L. (1992) Site directed mutagenesis at the *exo* III motif of ø29 DNA polymerase, overlapping structural domains for the 3'-5' exonuclease and strand displacement activivties. **EMBO J.** 11 (11) 4227-37

Sogo, J.M., Garcia, J.A., Penalva, M.A. & Salas, M. (1982) Structure of protein containing replicative intermediates of *B.subtilis* phage ø29 DNA. **Virology** 116: 1-18

Sor, F. & Fukuhara, H. (1985) Structure of a linear plasmid of the yeast *K. lactis* compact organisation of the linear genome. **Curr. Genet.** 9: 147-55

Sor, F., Wesolowski, M. & Fukuhara, H. (1983) Inverted terminal repetitions of the two linear DNA associated with the killer character of the yeast *K. lactis* **Nucleic Acid Res.** 11 (15) 5037-44

Southern, E. (1975) Detection of specific sequences amoung DNA fragments seperated by gel electrophoresis. **J. Mol. Biol.** 98: 503-17

Spanjaard, R.A. & van Duin, J. (1988) Translation of the sequence AGG-AGG yields 50% ribosomal frameshifting. **Proc. Natl. Acad. Sci.** 85: 7967-71

Spicer, E.K., Rush, J., Fung, C., Krantz, L.J.R., Karam, J.D. & Konigsberg, W.H. (1988) Primary structure of T4 DNA polymerase. **J. Biol. Chem.** 263 (16) 7478-86

Springer, B.A. & Sligar, S.G. (1987) High level expression of sperm whale myoglobin in *E. coli*. **Proc. Natl. Acad. Sci.** 84: 8961-5

Sreekrishna, K., Webster, T.D. & Dickson, R.C. (1984) Transformation of *K. lactis* with the G418 resistance gene of Tn 903. **Gene** 28: 73-81

Stader, J.A. & Silhavy, T.J. (1990) Engineering *E. coli* to secrete heterologous gene products. **Meth. Enzym.**185: 166-87

Stam, J.C., Meijer, M. & Stuitje, A.R. (1986) Efficient isolation of the linear killer plasmid of *K. lactis*: Evidence for location and expression in the cytoplasm and characterisation of their terminally bound proteins. **Nucleic Acid Res.** 14 (17): 6871

Stark, M.J.R. (1987) Multicopy expression vectors carrying the *lac* repressor gene for regulated high level expression of genes in *E. coli*. **Gene** 51: 255-67

Stark, M.J.R. (1988) Resolution of sequence discrepencies in the ORF1 region of the *K. lactis* plasmid K1. **Nucleic Acid Res.** 16 (2) 771

Stark, M.J.R. & Boyd, A. (1986) The Killer toxin of *K. lactis*: charcterisation of the toxin subunits and identification of the the genes which encode them. **EMBO J.** 5(8): 1995-2002

Stark, M.J.R., Mileham, A.J., Romanos, M.A. & Boyd, A. (1984) Nucleotide sequence and transcription analysis of a linear DNA plasmid associated with the killer character of the yeast *K. lactis*. **Nucleic Acid Res.** 12 (15) 6011-30

Stark, M.J.R., Boyd, A., Mileham, A.J. & Romanos, M.A. (1990) The plasmid encoded killer system of *K. lactis*: a review. **Yeast** 6: 1-29

Stillman, B.W. (1983) The replication of adenovirus DNA with purified proteins. **Cell** 35:7-9

Stillman, B.W. (1989) Initiation of eukaryotic DNA replication *in vitro*. **Annu. Rev. Cell Biol.** 5: 197-245

Stillman, B.W. & Lewis, J.B. (1981) Identification of the gene and mRNA for the adenovirus precurser terminal protein. **Cell** 23: 497-508

Stillman, B.W., Tamanoi, F. & Mathews, M.U. (1982a) Purification of an adenovirus-coded DNA polymerase that is required for initiation of DNA replication. **Cell** 31:613-23

Stillman, B.W., Topp, W.C. & Engler, J.A. (1982b) Conserved sequences at the origin of adenovirus DNA replication. **J. Virol.** 44 (2) 530-7

Stow, N.D. (1982) The infectivity of adenovirus genomes lacking DNA sequences from their left hand termini. **Nucleic Acid Res.** 10 (17) 5105-19

Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. **Meth. Enzymol.** 185: 60-89

Stuiver, M.H. & van der Vliet, P.C. (1990) Adenovirus DNA BP forms a multimeric protein complex with ds DNA and enhances binding of NF I. **J. Virol.** 65 (1) 379-86

Stunnenberg, H.G., Lange, H., Philipson, L., Miltonburg, R.T.V. & van der Vliet, P.C. (1988) High expression of functional adenovirus DNA polymerase and precursor terminal protein using recombinant vaccinia virus. **Nucleic Acid Res.** 16 (6) 2431-44

Sugisaki, Y., Gunge, N., Sakaguchi, K., Yamasaki, M. & Tamura, G. (1985) Transfer of DNA killer plasmids from *K. lactis* to *K. fragilis* and *Candida pseudotropicalis*. **J. Bacteriol.** 164 (3) 1373-5

Sussenbach, J.S. & van der Vliet, P.C. (1983) The mechanism of adenovirus DNA replication and the characterisation of replication proteins. **Curr. Top. Microbiol. Immun.** 109:53-70

Sutcliffe, J.G., Shinnick, T.M., Green, N. & Lerner, R.A. (1983) Antibodies that react with predetermined sites on proteins. **Science** 219: 660-6

Sutcliffe, J.G., Shinnick, T.M., Green, N., Liu, F.T., Niman, H.L. & Lerner, R.A. (1980) Chemical synthesis of a polypeptide predicted from nucleotide sequence allows detection of a new retroviral gene product. **Nature** 287: 801-5

Szakiel, G. & Pawlita, M. (1991) Inhibition of HIV-1 replication in human T-cells stably expressing antisense RNA. **J. Virol.** 65 (1) 468-72

Tainer, J.A., Getzoff, E.D., Alexander, H., Houghten, R.A., Olsen, A.J., Lerner, R.A. & Hendrickson, W.A. (1984) The reactivity of antipeptide antibodies is a function of the atomic mobility of sites in a protein. **Nature** 312: 127-33

Tam, J.P. (1988) Synthetic peptide design: Synthesis and properties of a high density antigenic peptide system. **Proc. Natl. Acad. Sci.** 85: 5409-13

Tam, J.P. & Zavala, F. (1989) Multiple antigen peptide. **J. Immunolog. Meth.** 124: 53-61

Tam, J.P., Clavijo, P., Lu, Y-A., Nussenzweig, V., Nussenzweig, R. & Zavala, F. (1990) Incorporation of T and B epitopes of the circumsporozoite protein in a chemically defined synthetic vaccine against maleria. **J. Exp. Med.** 171: 299-306

Tamanoi, F. & Stillman, B.W. (1982) Function of adenovirus terminal protein in the initiation of DNA replication. **Proc. Natl. Acad. Sci.** 79: 2221-5

Tamanoi, F. & Stillman, B.W. (1983a) Initiation of adenovirus DNA replication *in vitro* requires a specific DNA sequence. **Proc. Natl. Acad. Sci.** 80: 6446-50

Tamanoi, F. & Stillman, B.W. (1983b) The origin of adenovirus DNA replication. **Curr. Top. Microbiol. Immun.** 109:53-70

Tanguy-Rougeau, C., Chen, X.J., Wesolowski-Louvel, M. & Fukuhara, H (1990) Expression of a foreign kanamycin resistance gene in the linear killer DNA plasmids in yeast. **Gene** 91: 43-50

Temperley, S.M. & Hay, R.T. (1992) Recognition of Ad-2 origin of DNA replication by the virally encoded DNA polymerase and pTP. **EMBO J.** 11 (2) 761-8

Thompson-Jager, S. & Domdey, H. (1990) The intron of the yeast actin gene contains the promoter for an antisense RNA. **Curr. Genet.** 17: 269-73

Tokano, H., Kawano, S. & Kuroiwa, T. (1991) Telomeric structures in a linear mitochondrial plasmid from *Physarum polycephalum*. **Curr. Genet.** 20: 315-7

Tokunaga, M., Wada, N. & Hishinuma, F. (1987) Expression and identification of immunity determinants on linear DNA killer plasmids pGKL1 and pGKL2 in *K. lactis*. **Nucleic Acid Res.** 15: 1031-46

Tokunaga, M., Kawamura, A. & Hishinuma, F. (1989) Expression of pGKL1 killer 28K subunit in *S. cerevisiae*: identification of 28K subunit as a killer protein. **Nucleic Acid Res.** 17 (9) 3435-46

Tolun, A., Alestrom, P. & Pettersson, U. (1979) Sequence of ITR from different adenoviruses: Demonstration of conserved sequences and homology between SA7 and SV40 DNA. **Cell** 17: 705-13

Tomalski, M.D., Wu, J. & Miller, L.K. (1988) The location, sequence, transcription and regulation of a baculovirus DNA polymerase gene. **Virology** 167: 591-600

Tomizawa, J.I. & Itoh, T. (1981a) Plasmid ColE1 incompatability determined by interaction of RNA I with primer transcript. **Proc. Natl. Acad. Sci.** 78 (10) 6096-6100

Tomizawa, J.I., Itoh, T., Selzer, G. & Som, T. (1981b) Inhibition of ColE1 RNA primer formation by a plasmid specific small RNA. **Proc. Natl. Acad. Sci.** 78 (3) 1421-5

Tommasino, M. (1991) Killer system of *K. lactis*: the ORF 10 of pGKL2 plasmid encodes a putative DNA binding protein. **Yeast** 7: 245-52

Tommasino, M., Ricci, S. & Galeotti, C.L. (1988) Genome organisation of the killer plasmid pGKL2 from *K. lactis*. **Nucleic Acid Res.** 16 (13) 5863-78

Toulme, J.J. & Helene, C. (1988) Antimessenger oligodeoxyribonucleotides: an alternative to antisense RNA for artificial regulation of gene expression: A review. **Gene** 72: 51-8

Traynor, P.L. & Levings, C.S. (1986) Transcription of the maize S-2 mitochondrial plasmid. **Plant Mol. Biol.** 7: 255-63

Turpin, T., Garger, S.J., Marks, M.D. & Grill, L.K. (1987) Molecular cloning and physical characterisation of a *Brassica* linear mitochondrial plasmid. **Mol. Gen. Genet.** 209: 227-33

Uemori, T., Ishino, Y., Toh, H., Asada, K & Kato, I. (1993) Organisation and nucleotide sequence of the DNA polymerase gene from the archeon *Pyrococcus furiosus*. **Nucleic Acid Res.** 21 (2) 259-65

Uhlen, M. & Moks, T. (1990) Gene fusions for purpose of expression: An introduction. **Meth. Enzymol.** 185: 129-143

Varenne, S. & Lazdunski, C. (1986) Effect of distribution of unfavourable codons on the maximum rate of gene expression by a heterologous organism. **J. Theor. Biol.** 20: 99-110

Vierula, P.J., Cheng, C.K., Carst, D.A., Humphrey, R.W., Thomas, D.Y. & Bertrand, H. (1990) The kalilo senescence plasmid of *Neurospora intermedia* has covalently linked 5' terminal proteins. **Curr. Genet.** 17: 195-201

Volkert, F.C., Wilson, D.W. & Broach, J.R. (1989) Deooxyribonucleic acid plasmids in yeasts. **Micro. Revs** 53 (3) 299-317

Vliet, P.C.V., Keegstra, W. & Jansz, H.S. (1978) Complex formation between the Ad-5 DBP and ssDNA. **Eur. J. Biochem.** 86: 389-398

Vries, E., Driel, W.V., Tromp, J.V. & van der Vliet, P.C. (1985) Adenovirus DNA replication *in vitro*: site directed mutagenesis of the NF1 binding site of the Ad-2 origin. **Nucleic Acid Res.** 13 (13) 4935-52

Wang, C.Y., Looney, D.J., Li, M.L., Walfield, A.M., Hosein, J.Y.B., Tam, J.P. & Wang-Stall, F. (1991) Long term high titer neutralising activity induced by octomeric synthetic HIV-1 antigen. **Science** 254: 285-8

Wang, G-H. & Seeger, C. (1992) The reverse transcriptase of Hepatitis B virus acts as a protein primer for viral DNA synthesis. **Cell** 71 (13) 663-70

Wang, K. & Pearson, G.D. (1985) Adenovirus sequence required for replication *in vitro*. **Nucleic Acid Res.** 13 (14) 5173-87

Wang, T.S-F. (1991) Eukaryotic DNA polymerases. **Ann. Rev. Biochem.** 60: 513-52

Wang, T.S.F., Wong, S.W. & Korn, D. (1989) Human DNA polymerase-α: Predicted functional domains and relationships with viral DNA polymerases. **Fed. Am. Soc. Exp. Biol. J.** 3 (1) 14-21

Watabe, K. & Ito, J. (1983) A novel DNA polymerase induced by *B.subtilis* phage ø29. **Nucleic Acid Res.** 11 (23) 8333-9

Watabe, K., Shih, M-F., Sugino, A. & Ito, J. (1982) *In vitro* replication of bacteriophage ø29 DNA. **Proc. Natl. Acad. Sci.** 79: 5245-8

Watabe, K., Leusch, M. & Ito, J. (1984a) Replication of bacteriophage ø29 DNA *in vitro*: The roles of terminal protein and DNA polymerase. **Proc. Natl. Acad. Sci.** 81: 5374-8

Watabe, K., Leusch, M.S. & Ito, J. (1984b) A 3'-5' exonuclease activity is associated with phage ø29 DNA polymerase. **Bioch. Biophy. Res. Com.** 123 (3) 1019-26

Watson, C.J. & Hay, R.T. (1990) Expression of adenovirus type 2 DNA polymerase in insect cells infected with a recombinant baculovirus. **Nucleic Acid Res.** 18 (5) 1167-73

Weintraub, H.M. (1990) Antisense RNA and DNA. **Sci. Am.** 262: 34-40

Weintraub, H.M., Izant, J.G. & Harland, R.M. (1985) Antisense RNA as a molecular tool for genetic analysis. **Trends In Genet.** 1 (1) 23-5

Wesolowski, M. & Fukuhara, H. (1981) Linear mitochondrial deoxyribonucleic acid from the yeast *Hansenula mrakii*. **Mol.Cell.Biol.** 1 (6) 387-93

Wesolowski-Louvel, M. & Fukuhara, H (1990) A palindromic mutation of the linear killer plasmid K2 of yeast. **Nucleic Acid Res.** 18 (16) 4877-82

Wesolowski, M., Algen, A. & Fukuhara, H. (1982a) Killer DNA plasmids of the yeast *K. lactis*: III plasmid recombination. **Curr. Genet.** 5: 205-8

Wesolowski, M., Goffrini, P. & Fukuhara, H. (1982b) Killer DNA plasmids of the yeast *K. lactis*: I. Mutations affecting killer phenotype. **Curr. Genet.** 5: 191-7

Wesolowski-Louvel, M., Tanguy-Rougeau, C. & Fukuhara, H. (1988) A nuclear encoded gene required for the expression of the linear DNA associated killer system in the yeast *K. lactis*. **Yeast** 4: 71-81

Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. & Van Regenmortel, M.H.V. (1984) Correlation between segmental mobility and the location of antigenic determinants in proteins. **Nature** 311: 123-6

Wides, R.J., Challberg, M.D., Rawlins, D.R. & Kelly, T.J. (1987) Adenovirus origin of DNA replication: sequence requirements for replication *in vitro*. **Mol. Cell. Biol.** 7 (2) 864-74

Wilkinson, D.L. & Harrison, R.C. (1991) Predicting the solubility of recombinant proteins in *E. coli*. **Biotechnology** 9: 443-8

Wilson, D.W. & Meacock, P.A. (1988) Extranuclear gene expression in yeast: evidence for a plasmid encoded RNA polymerase of unique structure. **Nucleic Acid Res.** 16 (16) 8097-8112

Wong, S.W., Wahl, A.F., Yuan, P-M., Arai, N., Pearson, B.E., Arai, K., Korn, D., Hunkapiller, M.W. & Wang, T.S-F. (1988) Human DNA polymerase-α gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. **EMBO J** 7 (1) 37-47

Worsham, P.C. & Bolen. (1990) Killer toxin produced in *Pichia acaciae* is associated with linear DNA plasmids. **Curr. Genet.** 18: 389-93

Yanofsky, S., Kawamura, F., & Ito, J. (1976) Thermolabile transfecting DNA from ts mutant of phage ø29. **Nature** 259: 60-3

Yehle, C.O. (1978) Genome linked protein associated with the 5' termini of bacteriophage ø29 DNA. **J. Virol.** 27 (3) 776-83

Yoshikawa, H. & Ito, J. (1981) Terminal proteins and short ITR of the small *Bacillus* bacteriophage genomes. **Proc. Natl. Acad. Sci.** 78 (4) 2596-2600

Yui, Y., Katayose, Y. & Shishido, K. (1988) Two linear plasmid like DNA elements simultaneously maintained in *Plurotus astreatus*. **Biochim. Biophys. Acta.** 951: 53-60

Zaballos, A., Mellado, R.P. & Sallas, M. (1986) Initiation of phage ø29 DNA replication by mutants with deletions at the N-terminal end of the terminal protein. **Gene** 43: 103-10

Zaballos, A. & Sallas, M. (1989a) Functional domains in the bacteriophage ø29 terminal protein for interaction with ø29 DNA polymerase and with DNA. **Nucleic Acid Res.** 17 (24) 10353-65

Zaballos, A., Lazaro, J.M., Mendez, E., Mellado, R.F. & Salas, M. (1989b) Effects of internal deletions on the priming activity of the phage ø29 TP. **Gene** 83: 187-95

Zang, S., Zubay, G. & Goldman, E. (1991) Low usage codons in *E. coli*, yeast, fruit flies and primates. **Gene** 105: 61-72

Appendix 1.

Plasmid Constructions Made During the Course of this Study.

plasmid name	plasmid size (kb)	plasmid structure
pCA1	7.68	3 kb *Eco*R I-*Sal* I K1 ORF 1 (PCR) in pTTQ8
pCA2	7.68	3 kb *Eco*R I-*Sal* I K2 ORF 2 (PCR) in pTTQ8
pCA3	7.25	3 kb *Eco*R I-*Sal* I K1 ORF 1 (from pCA1) in pRIT2T
pCA4	7.25	3 kb *Eco*R I-*Sal* I K2 ORF 2 (from pCA2) in pRIT2T
pCA5	4.4	1.7 kb *Pst* I-*Sal* I K1 ORF 1 from pCA1 in pIC19R
pCA6	6.6	1.7 kb K1 ORF 1 *Bam*H I-*Eco*RI (from pCA9) in pGEX2T
pCA7	6.9	2 kb K2 ORF 2 (*Bgl*II-*Eco*R1) in pGEX1
pCA8	4.5	*Eco*RV deletion of pET3a
pCA9	7.5	3 kb *Nhe* I-*Bam*H I K2 ORF 2 (PCR) in pCA8
pCA10	5.42	470 bp *Bgl*II-*Sau*3a K2 ORF 2 in pGEX1
pCA11	5.8	900 bp *Sau*3a-*Eco*R I K2 ORF 2 in pGEX1
pCA12	5.3	390 bp *Bam*H I-*Eco*R I (PCR) K2 ORF 2 in pGEX2T
pCA13	4.9	390 bp *Nhe* I-*Bam*H I (PCR) K2 ORF 2 in pCA8
pCA14	5.42	470 bp *Bgl*II-*Sau*3a K2 ORF 2 in pGEX1 (in wrong orientation-opposite to pCA10)
pCA15		2.5 kb *Bgl*II-*Sma*1 *trp1* fragment of pG1 in *Bam*H I-*Nru* I cut KEp6
pCA16	3.1	250 bp PCR antisense fragment of K1 ORF 1 N-terminal in *Hph* I cut pCR1000
pCA17		250 bp *Bam*H1-*Sal*1 antisense fragment (PCR) from pCA16 in pCA15
pCA17		1.7 kb *Pst* I-*Sal* I PCR fragment of K1 ORF 1 in *Pst* I-*Sal* I cut pAG111 (Maxwell)
pCA18	2.7kb	*Bam*H I-*Sal* I 9E10 epitope tag in pIC19H
pCA19	3.3kb	0.6 kb *Pst* I Km^R gene fragment from pEMBL-4K in pCA18
pCA20	3.3	0.6 kb *Pst* I Km^R gene from pEMBL-4K in pIC20R containing poly-histidine tag (Brophy)
pCA21	5	1.7 kb *Pst* I-*Sal* I fragment of K1 ORF 1 in pCA20 containing poly-his tag
pCA22	5	1.7 kb *Pst* I-*Sal* I K1 ORF1 fragment in pCA19
pCA23	8.3	3.3 kb *Nde* I fragment of pMS11 (Soond) containing *trp1* gene + flanking K1 sequences in pCA22
pCA24	3.45	250 bp *Bam*H I-*Sal* I antisense fragment from pCA16 in pBSM13+
pCA25	5	1.7 kb of K1 ORF 1 (PCR) in pBSM13+ *Sma*I site
pCA26	4.4	1.7 kb *Pst* I-*Sal* I K1 ORF1 fragment from pCA25 in *Pst*I-*Sal*I site of pCA20
pCA27	7.7	3.3 kb *Nde* I (filled in) fragment from pMS11 in *Sma* I site of pCA26
pCA28	2.86	160 bp *Bam*H I-*Eco*R I synthetic gene fragment cloned into *Bam*H I-*Eco*R I site of pIC20H
pCA29	4.66	160 bp *Bam*H I-*Bgl* II synthetic gene fragment from pCA28 cloned into *Bam*H I site of pET3a
pCA30	4.46	160 bp *Bam*H I-*Eco*R I synthetic gene fragment from pCA28 cloned into *Bam*H I-*Eco*R I site of pGEX2T
pCA31		*Eco*R I-*Bam*H I SP2 epitope cloned into *Bam*H I-*Eco*R I site of pGEX2T

APPENDIX 2.

PCR primer oligonucleotides used during this study.

Primer	Length (nt)	Sequence (5' to 3')	Used with:	Integral restriction site	Target gene	Target sequence in plasmid	Fragment amplified (Kb)	Target vector
CA1	21	AAA ATG AAT TCC AAA GAT AAG	CA2	*Eco*R I	K1 ORF 1	234-254	3	pTTQ18
CA2	25	TTG TCG ACT TAA GTA GCT TTC ACG G	CA1	*Sal* I	K1 ORF 1	3183-3198	3	pTTQ18
CA3	21	AAA ATG AAT TCT GAA TTA GCA	CA4	*Eco*R I	K2 ORF 2	3868-3898	3	pTTQ18
CA4	26	TTG TCG ACT TAA GTC TGT TTT ATA TC	CA3	*Sal* I	K2 ORF 2	901-918	3	pTTQ18
CA5	26	TTG CTA GCA TGA ATG ATG AAT TAG CA	CA6	*Nhe* I	K2 ORF 2	3869-3885	3	pCA8
CA6	26	TTG GAT CCT TAA GTC TGT TTT ATA TC	CA5	*Bam*H I	K2 ORF 2	901-918	3	pCA8
CA7	24	CC GTC GAC AGT AGC TTT CAC GGT C	CA1	*Sal* I	K1 ORF 1	3182-3195	3	pBSM13$^+$
CA8	25	CCC GGA TCC AAT GTT TCC GCA TTT T	CA5	*Bam*H I	K2 ORF 2	3496-3512	0.39	pCA8
CA9	27	CCC GGA TCC ATG AAT GAT GAA TTA GCA	CA10	*Bam*H I	K2 ORF 2	3868-3885	0.39	pGEX2T
CA10	25	CCC GAA TTC AAT GTT TCC GCA TTT T	CA9	*Eco*R I	K2 ORF 2	3496-3512	0.39	pGEX2T
CA11	25	CCG TCG ACA AAA ATG GAT TAC AAA G	CA12	*Sal* I	K1 ORF 1	233-249	0.345	pCA15
CA12	23	(5') CCG GAT CCA TTG TCT CTC GTC TA	CA11	*Bam*H I	K1 ORF 1	260-274	0.345	pCA15

APPENDIX 2B.

The oligonucleotides used for cloning during this study.

Oligo	Length (nt)	Sequence (5' to 3')	Function	Integral restriction sites	Target plasmid
CA13	39	TC GAC GAA CAA AAA TTA ATT TCT GAA GAA GAT TTA TAA G	Anneals to CA14 to form the top strand of the 9E10epitope tag for incorporation into the recombinant K1 ORF 1.	*Sal* I	pIC
CA14	39	GA TCC TTA TAA ATC TTC TTC AGA AAT TAA TTT TTG TTC G	Anneals to CA13 to form the bottom strand of the 9E10 epitope tag	*Bam*H I	pIC
CA15	27	TC GAC CAT CAT CAT CAT CAT CAT TAA G	Anneals to CA16 to form the top strand of the poly-histidine tag for incorporation into the recombinant K1 ORF 1.	*Sal* I	pIC
CA16	27	GA TCC TTA ATG ATG ATG ATG ATG ATG G	Anneals to CA16 to form the bottom strand of the poly-histidine tag	*Bam*H I	pIC
CA17	98	CC CCC GGA TCC ATG AAC GAC GAA CTG GCT TTC CTG AAC TCT CAG ATC GAC GAC TAC TCT GAC GAC ATC GAG TTC CTG CGT ACC GAA GTT TTC CGT ATC	Anneals to CA18 to form the top strand of the resynthesised N-terminal 168 bp of K2 ORF 2.	*Bam*H I	pET3a & pGEX2T
CA18	97	CCC CCG AAT TCG ATG GTT TCT TTG ATA GCT TCT TTA GAA GAG TTT TTC ATC AGA GAG TAG ATA GAT TTA AGA GTA ACG GTT GAT ACG GAA AAC TTC GG	Anneals to CA17 to form the bottom strand of the resynthesised N-terminal 168 bp of K2 ORF 2.	*Eco*R I	pET3a & pGEX2T
CA19	30	GAT CCA CCT ACG GTG ACA TCA ACA AAG ACG	Anneals to CA20 to form the top strand of the potential epitope corresponding to the second synthetic peptides (SP2)	*Bam*H I	pGEX2T
CA20	30	AAT TCG TCT TTG TTG ATG TCA CCG TAG GTG	Anneals to CA19 to form the bottom strand of the potential epitope corresponding to the second synthetic peptide (SP2)	*Eco*R I	pGEX2T