CLONING OF THE T4 POLYNUCLEOTIDE KINASE GENE AND AMPLIFICATION OF ITS PRODUCT

by

CAROL ANNE MIDGLEY

A thesis presented for the degree of Doctor of Philosophy

Department of Molecular Biology
University of Edinburgh

October 1985

TO MY MOTHER AND FATHER

FOREWORD

The work presented and the composition of this thesis has been my own. Many of the approaches to the work were devised in collaboration with my supervisor, Dr. Noreen E. Murray.

Carol Anne Midgley
Department of Molecular Biology,
University of Edinburgh.

ACKNOWLEDGEMENTS

I would like to thank everyone who has helped me towards completing this thesis, particularly Ken Murray and David Finnegan for their advice and encouragement, Sandra Bruce, Heather Houston, Jean Manson, Barbara Will and many others who have spent time teaching me how to do things. Thanks to Keith Rand and the MRC Laboratory of Molecular Biology, Cambridge, for initial help with sequencing techniques, and to Andrew Coulson, John Collins, Lindsay Sawyer and Arthur Robinson for advice on computing. Thanks to everyone in the Department, particularly Noreen's group namely, Anne, Annette, Frances, Wil and Gill for various supporting roles, and thanks to the ladies who make media and who do the washing up, particularly Marion and Helen. Thanks to Kathie Harris for typing the thesis, Annie Wilson for some of the figures, Graham Brown for photography and David Meek for proof reading.

Much love and thanks to my parents for their support and encouragement, and finally, many thanks to my supervisor, Noreen Murray, for proof reading this thesis several times and for three years of patience and good advice.

ABSTRACT

The T4 gene (*pseT*) encoding polynucleotide kinase (pnk) has been cloned in λ vectors on a single T4 *Eco*RI fragment (fragment 21) and expression of active pnk has been achieved. Induction of a λ*E*⁻*W*⁻*S*⁻*c*I857 prophage in which the *pseT* gene can be transcribed from the late λ promoter, $p_{R'}$, leads to more than 100 fold amplification of pnk activity in relation to infection with T4 wild-type phage. The pnk polypeptide constitutes about 7% of the total soluble cell protein in the induced lysogen. The purified enzyme, as expected, is both a 5'-kinase and a 3'-phosphatase. The amino acid sequence of the ORF encoding pnk was deduced from the DNA sequence of fragment 21, both for the active pnk polypeptide and for an inactive pnk polypeptide encoded by a deletion mutant including fragment 21, which apparently induced neither 5'-kinase or 3'-phosphatase activity. The latter polypeptide lacked a Pro-Gly dipeptide within a part of the sequence which shared homology with the adenine nucleotide binding sites of several proteins. The dipeptide is within the region corresponding in adenylate kinase to a flexible loop thought to take part in a conformational change after AMP binds to the enzyme. The loss of amino acids from a similar loop may prevent the defective pnk enzyme from binding adenine nucleotide substrates, or may constrict a conformational change and prevent the enzyme from achieving an activated state. A second region

of the pnk sequence shares homology with phosphoglycerate kinase, yeast inorganic pyrophosphatase, and histone 2b from various species.

ABBREVIATIONS

bp	-	base pair
DE	-	delayed early (gene)
g30	-	gene 30
gp30	-	protein product of gene 30
HMC	-	hydroxymethylcytosine
IE	-	immediate early (gene)
IPTG	-	isopropyl-β-D-thiogalactoside
kb	-	kilobase
kd	-	kilodalton
M_r	-	molecular weight
moi	-	multiplicity of infection
ORF	-	open reading frame
pfu	-	plaque forming unit
pnk	-	polynucleotide kinase
rbs	-	ribosome binding site
SDS-PAGE	-	sodium dodecyl sulphate polyacrylamide gel electrophoresis
TEMED	-	N,N,N',N'-tetramethylethylenediamine
UV	-	ultraviolet light
Xg	-	5-bromo-4-chloro-3 indolyl-β-D-galactoside
AMPS	-	Ammonium persulphate
BSA	-	Bovine serum albumin
DTT	-	Dithiothreitol
MES	-	2-[N-Morpholino]ethanesulphonic acid
SSC	-	Standard saline citrate

CONTENTS

2.2 Methods (contd.)

2.2 Methods (contd.)

CHAPTER 1

INTRODUCTION

1.1. BACTERIOPHAGE T4

Bacteriophage T4 is a complex DNA virus with a total genome length (Figure 1.1) of 166,000 nucleotide pairs (Kim and Davidson, 1974). About 140 genes have now been identified (Wood and Revel, 1976) and many are functionally characterised.

The phage consists of particles or virions each made up of a protein coat or capsid that surrounds and protects the viral genome. T4 is a member of the class of urophages which are characterised by a large head filled with double-stranded DNA and a tail through which the DNA is extruded during infection (Figure 1.2). An unusual feature of T4 DNA is that its genetic map is circular even though the DNA of each virion is linear (Streisinger et al., 1964). Mature phage DNA also displays about 3% terminal redundancy (Thomas and Rubenstein, 1964). Since T-even phage package their DNA from concatemeric DNA molecules by a headful mechanism (Ritchie and White, 1972) rather than by cutting at specific sites, the above features ensure that a whole genome is packaged. T4 compensates for deletions of non-essential regions of the genome by increasing the length of the terminal repeats.

The DNA of T-even phages contains hydroxymethylcytosine (HMC) instead of cytosine (Wyatt and Cohen, 1953). These residues are glucosylated by the α or β glucosyl transferases (Lehmann and Pratt, 1960). In addition some adenine residues are methylated (Hattman, 1970).

Figure 1.1

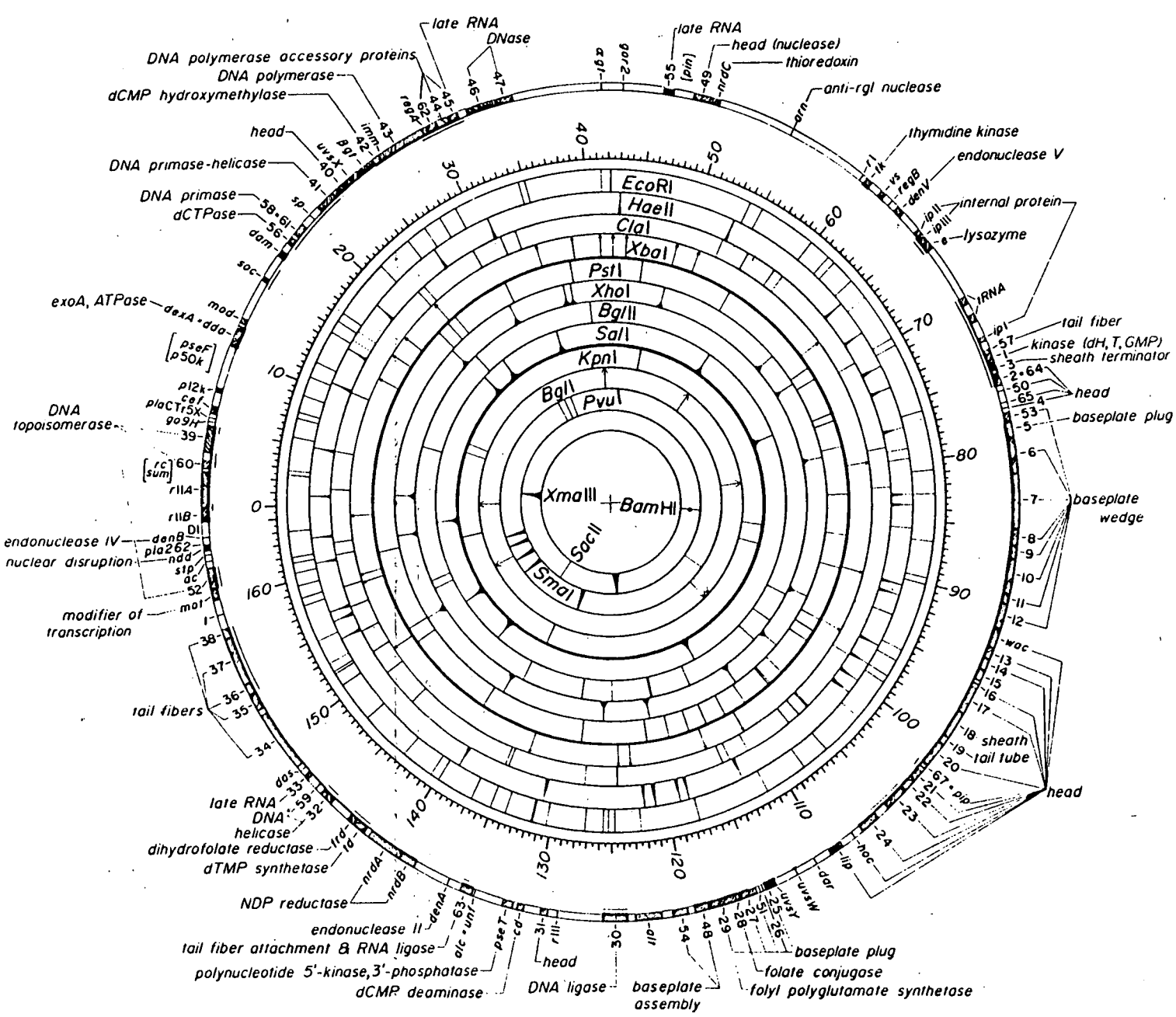

late RNA
DNase
late RNA
head (nuclease)
thioredoxin
anti-rgl nuclease
thymidine kinase
endonuclease V
internal protein
lysozyme
tRNA
tail fiber
kinase (dH, T, GMP)
sheath terminator
head
baseplate plug
baseplate wedge
head
sheath
tail tube
baseplate plug
folate conjugase
folyl polyglutamate synthetase
baseplate assembly
DNA ligase
head
endonuclease II
tail fiber attachment & RNA ligase
polynucleotide 5'-kinase, 3'-phosphatase
dCMP deaminase
NDP reductase
dTMP synthetase
dihydrofolate reductase
DNA helicase
late RNA
tail fibers
modifier of transcription
endonuclease IV
nuclear disruption
DNA topoisomerase
exoA, ATPase
dCTPase
DNA primase
DNA primase-helicase
head
dCMP hydroxymethylase
DNA polymerase
DNA polymerase accessory proteins
EcoRI
HaeII
ClaI
XbaI
PstI
XhoI
BglII
SalI
KpnI
BglI
PvuI
XmaIII
BamHI
SacII
SmaI

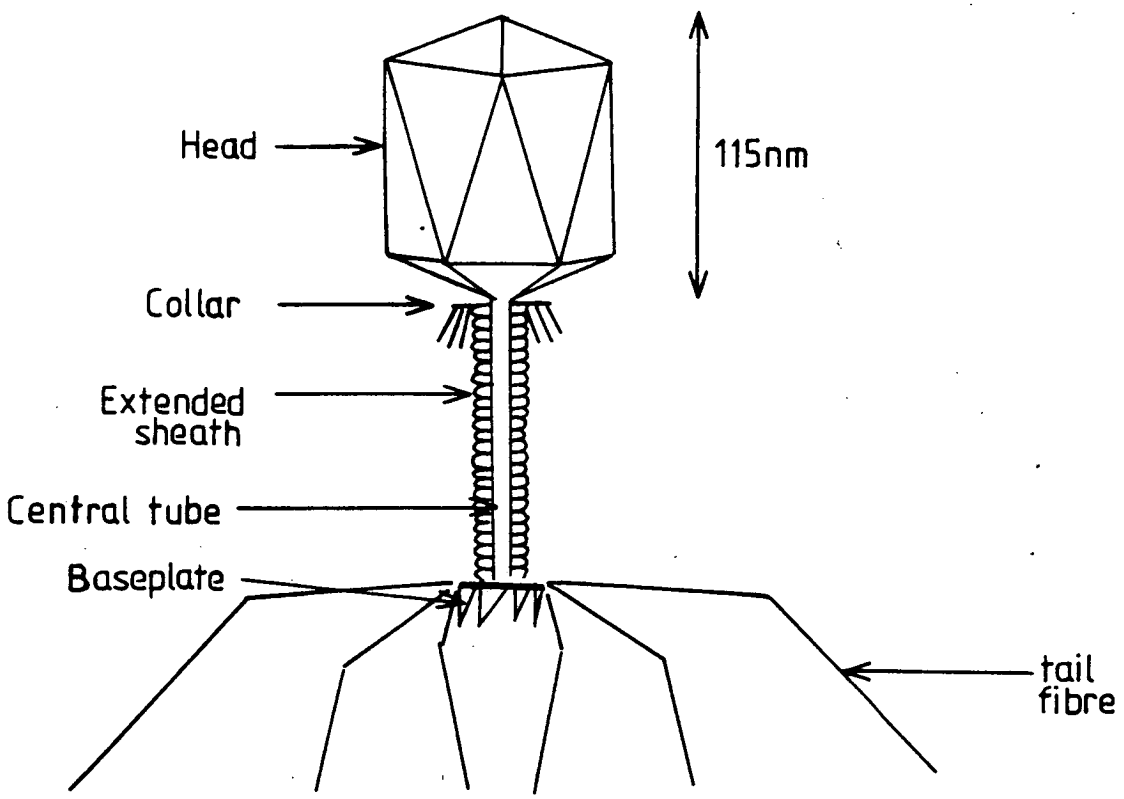
Figure 1.1

The genetic map of bacteriophage T4 correlated with the restriction map, from Kutter and Rüger (1983). Regions indicated by shading in the outer circle are where restriction mapping has been used to help locate genes.

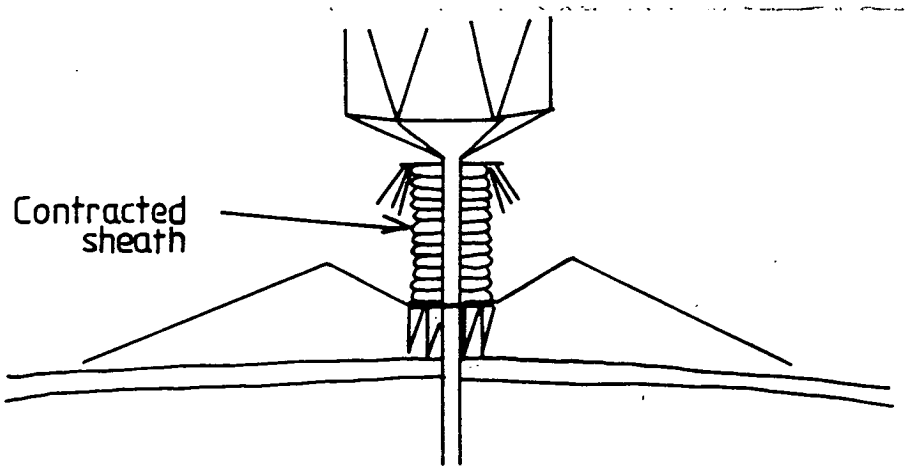
The genes of T4 can be divided according to their function into two main classes: cell metabolism and phage particle assembly (Figure 1.3). In addition they are classified by the timing of appearance of their products as immediate-early (IE), delayed-early (DE), quasi-late (Q) or true-late (L), (O'Farrell and Gold, 1973). The direction of transcription has been determined for a number of genes, and evidence from mRNA hybridisation to separated T4 DNA strands indicates that probably all early and quasi-late genes are transcribed in an anticlockwise direction in relation to the T4 map (Figure 1.1), whereas late genes are transcribed in a clockwise direction (Guha et al., 1971). Most metabolic functions are controlled by genes with early or quasi-late characteristics while most assembly functions are controlled by late genes (Wood and Revel, 1976).

Essential genes (designated mainly by numbers or lower case single letters) have been identified by conditionally lethal mutations (Epstein et al., 1963) and include 22 of the metabolic functions and all but 10 of the assembly functions. Those genes non-essential on normal laboratory E.coli strains are designated by two or three letter symbols. Most of these are in the metabolic category and many seem to augment the phage burst size e.g. by providing more abundant substrates for DNA replication,or by helping the phage to deal with the host's intracellular environment. However, many of their functions are unknown and some may in fact perform

Figure 1.2


Head
115nm
Collar
Extended sheath
Central tube
Baseplate
tail fibre

(a)


Contracted sheath

(b)

Figure 1.2

A diagram of bacteriophage T4 showing the major structural components.

essential functions which can be carried out by alternative host pathways. Comparison of Figures 1.1 and 1.3 show that T4 genes exhibit considerable clustering according to function. This may be a result of selective pressure since it minimises recombination between genes encoding proteins that must interact structurally (Stahl and Murray, 1966). On the basis of identified genes, 15% of the genome is used to code for essential metabolic functions, 39% for non-essential functions and 36% for particle assembly (Wood and Revel, 1976).

1.2 STAGES IN INFECTION OF E.COLI BY BACTERIOPHAGE T4

When T4 phage infect a cell only DNA and a few auxiliary proteins are injected (Hershey and Chase, 1952), so for the first few minutes there are no complete infective particles. After this initial eclipse period, phage particles appear and increase steadily in number until each cell contains 200-400 phage. The host cell eventually bursts releasing them.

Details of phage structure and multiplication, which have been elucidated mainly through conditional mutants (Epstein et al., 1963) are described below. Once inside the cell, the viral genome stops host DNA synthesis and gene expression and eventually the host genome is destroyed. The cell is instructed to make new virions by the T4 genome which encodes both the enzymes for DNA synthesis and the capsid proteins.

Figure 1.3

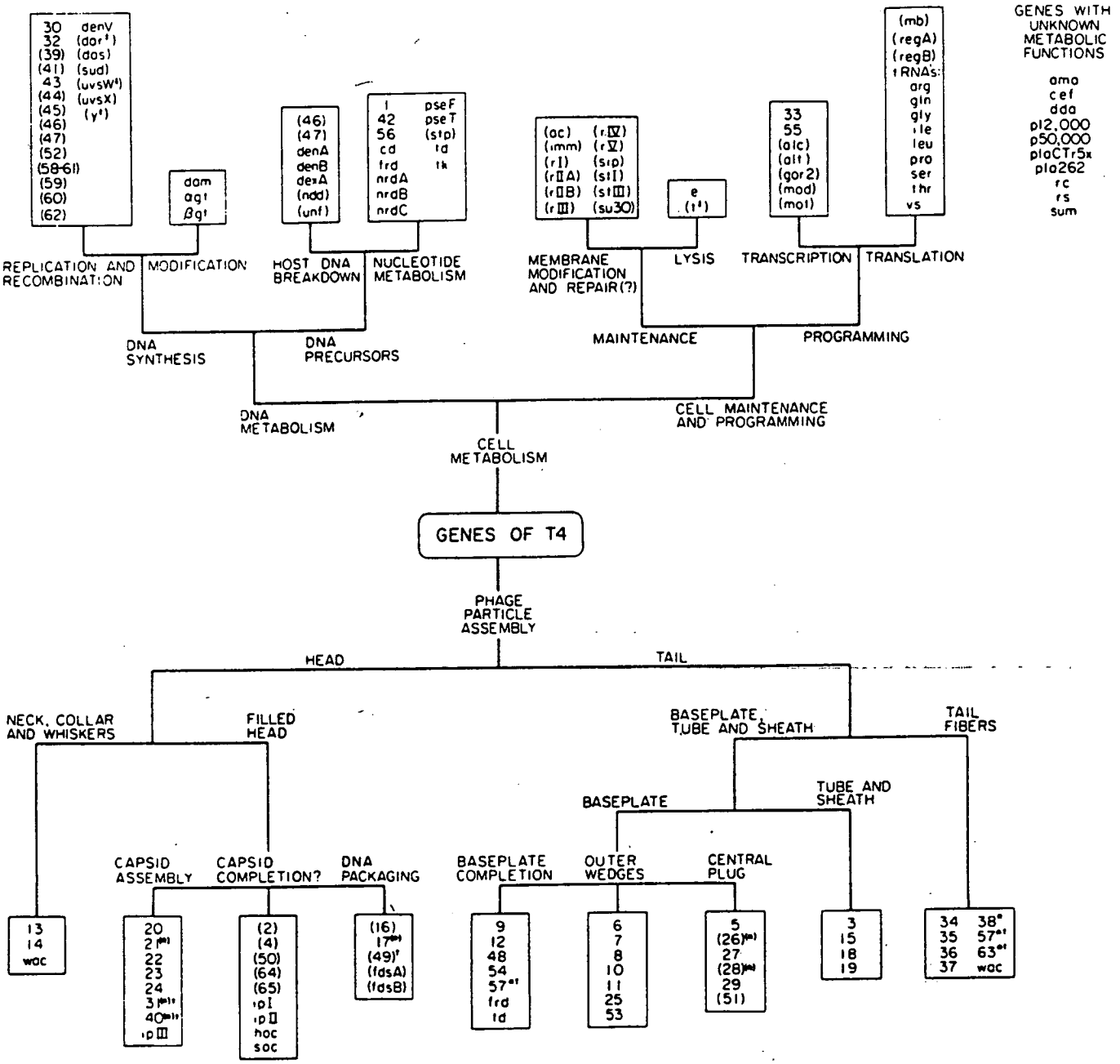

30 denV
32 (dor⁻)
(39) (dos)
(41) (sud)
43 (uvsW⁺)
(44) (uvsX)
(45) (y⁺)
(46)
(47)
(52)
(58-61)
(59)
(60)
(62)
dam
agt
βgt
(46)
(47)
denA
denB
dexA
(ndd)
(unf)
I pseF
42 pseT
56 (stp)
cd td
frd tk
nrdA
nrdB
nrdC
(ac) (rIX)
(imm) (rV)
(rI) (stp)
(rIIA) (stI)
(rIIB) (stIII)
(rIII) (su30)
e
(t⁺)
33
55
(alc)
(alt)
(gor2)
(mod)
(mot)
(mb)
(regA)
(regB)
tRNA's:
arg
gln
gly
ile
leu
pro
ser
thr
vs
GENES WITH UNKNOWN METABOLIC FUNCTIONS
amo
cef
dda
p12,000
p50,000
plaCTr5x
pla262
rc
rs
sum
REPLICATION AND RECOMBINATION
MODIFICATION
HOST DNA BREAKDOWN
NUCLEOTIDE METABOLISM
MEMBRANE MODIFICATION AND REPAIR(?)
LYSIS
TRANSCRIPTION
TRANSLATION
DNA SYNTHESIS
DNA PRECURSORS
MAINTENANCE
PROGRAMMING
DNA METABOLISM
CELL MAINTENANCE AND PROGRAMMING
CELL METABOLISM
GENES OF T4
PHAGE PARTICLE ASSEMBLY
HEAD
TAIL
NECK, COLLAR AND WHISKERS
FILLED HEAD
BASEPLATE, TUBE AND SHEATH
TAIL FIBERS
BASEPLATE
TUBE AND SHEATH
CAPSID ASSEMBLY
CAPSID COMPLETION?
DNA PACKAGING
BASEPLATE COMPLETION
OUTER WEDGES
CENTRAL PLUG
13
14
wac
20
21
22
23
24
31
40
ipIII
(2)
(4)
(50)
(64)
(65)
ipI
ipII
hoc
soc
(16)
17
(49)
(fdsA)
(fdsB)
9
12
48
54
57
frd
td
6
7
8
10
11
25
53
5
(26)
27
(28)
29
(51)
3
15
18
19
34
35
36
37
38
57
63
wac

Figure 1.3

Functional classification of T4 genes, from Wood and Revel (1976). Genes whose functions are at least generally known are listed in boxes representing different functional categories. Some genes known to have dual functions are listed under two functional categories.

1.2.a. Attachment

The tail fibres (Figure 1.2) are the site of specific recognition for attachment to the host cell (Simon and Anderson, 1967). The host cell receptors are the lipopolysaccharide molecules of the outer cell wall (Wilson et al., 1970), however irreversible attachment and penetration occurs only at about 200 sites (Bayer, 1968). Once the phage is attached, the sheath contracts and drives the tail tube into the cell envelope.

The infecting phage DNA enters the cell accompanied by several minor proteins including gp2 and gp64 which are probably attached to the termini of DNA molecules thereby protecting them against bacterial nucleases (Silverstein and Goldberg, 1976). 2-5 min after infection, host DNA, RNA and protein synthesis virtually terminate and subsequent macromolecular synthesis is almost entirely phage mediated. Inhibition of host gene expression occurs at both the transcriptional and post-transcriptional levels, and although host DNA continues to be transcribed for several minutes there is an immediate block on induction of all inducible host specific enzymes e.g. β-galactosidase (Monod and Wollman, 1947).

T4 produces a series of nucleases which degrade cytosine-containing DNA after infection. The product of the gene denA is endonuclease II which nicks

double-stranded DNA at the 5' side of cytosines (Sadowski and Hurwitz, 1969). An exonuclease activity, probably host encoded, appears to enlarge the nick to a gap so that the phage gene product of denB, endonuclease IV, can cut to the 5' side of cytosines in the single-stranded region produced (Sadowski and Bakyta, 1972). These fragments are finally degraded to mononucleotides presumably by a phage encoded exonuclease. This process of host DNA degradation serves to avoid transcriptional competition and enrich the nucleotide pool.

1.2.b. DNA replication

The conversion of cytosine to HMC occurs at the mononucleotide level (Figure 1.4). HMC-DNA is synthesised without glucosylation, then phage encoded enzymes glucosylate HMC groups in the DNA (Revel and Luria, 1970).

Replication of T-even phage DNA is semi-conservative and dispersive i.e. segments of the infecting DNA strands are distributed among several progeny (Kozinski, 1961). Circularisation of the genome does not seem to be necessary for successful replication (Kozinski and Doermann, 1975). If all of the required phage genes are functional, replication starts about 6 min after infection and occurs bi-directionally, probably from multiple origins (Delius et al., 1971; Howe, et al. 1973; Halpern et al., 1979). Initiation of leading strand

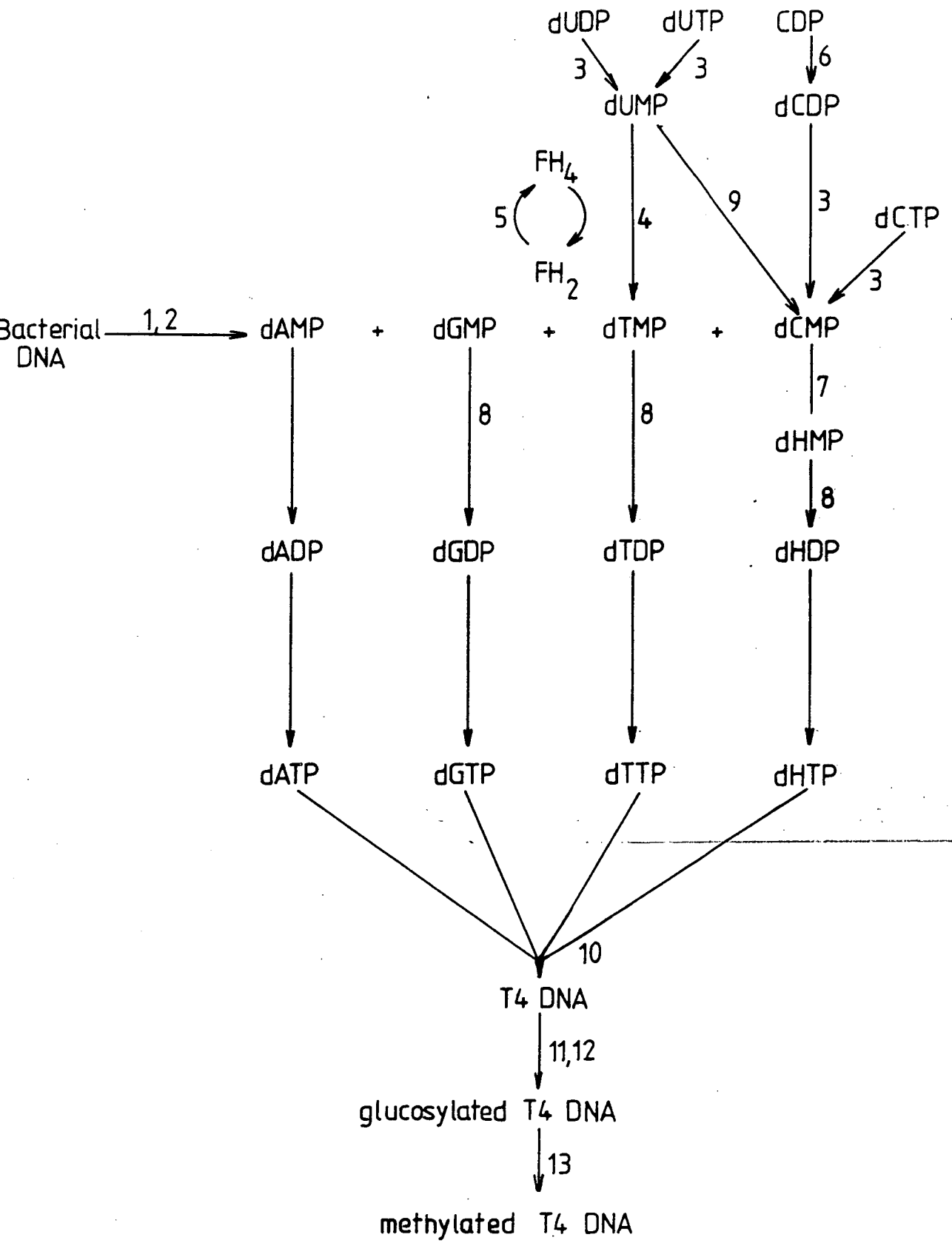

dUDP
dUTP
CDP
3
3
6
dUMP
dCDP
FH4
5
FH2
4
9
3
dCTP
3
Bacterial DNA
1,2
dAMP
+
dGMP
+
dTMP
+
dCMP
8
8
7
dHMP
8
dADP
dGDP
dTDP
dHDP
dATP
dGTP
dTTP
dHTP
10
T4 DNA
11,12
glucosylated T4 DNA
13
methylated T4 DNA

Figure 1.4

Figure 1.4

The enzymatic biosynthesis of T4 phage DNA. Reactions catalysed by virus coded functions are numbered and the relevant enzymes are listed below.

1. Endonuclease
2. Exonuclease
3. Deoxycytidine and deoxyuridine tri- and diphosphatase
4. Thymidylate synthetase
5. Dihydrofolate reductase
6. Cytidine diphosphate reductase
7. Deoxycytidylate hydroxymethylase
8. Deoxynucleotide kinase
9. Deoxycytidylate deaminase
10. DNA polymerase
11. α-Glucosyl transferase
12. β-Glucosyl transferase
13. DNA methylase

synthesis requires an RNA primer provided by E.coli RNA polymerase (Luder and Mosig, 1982). A T4 primase primes the Okazaki fragments of the lagging strand which are joined on to the growing strand by DNA ligase.

In vitro synthesis of DNA requires a number of purified phage encoded proteins. gp43 (the DNA polymerase) gp41/61, gp44/62, gp45 and gp32 together with the four NTPs, dNTPs and Mg^{++} replicate DNA at close to the in vivo rate (Alberts et al., 1975). gp44 and gp62 form a complex with ATPase activity which gp45 stimulates (Piperno et al., 1978; Mace and Alberts, 1984). gp32 is a helix-destabilising protein which associates with single-stranded DNA. These accessory proteins probably help the polymerase to overcome 'pausing' at hairpin structures in the DNA (Huang et al., 1981). gp41 and gp61 are part of the T4 primase (Liu and Alberts, 1981; Silver and Nossal, 1982).

The first round of replication is initiated de novo at origin sequences, but when replication reaches an end of the molecule, the template strand for lagging strand synthesis is single-stranded at the 3' end and this segment will be able to invade a homologous region of another molecule or the terminally redundant sequences of the same molecule (Danenberg and Mosig, 1983). These invasions form new recombinational forks eventually generating a

complex network of branched and looped T4 DNA (Altman and Lerman, 1970; Hamilton and Pettijohn, 1976). Reinitiation from origin sequences is rare, as modification of RNA polymerase during infection interferes not only with recognition of early promoters but also of replication origins (Rabussay, 1982). Most secondary initiations must start from recombinational intermediates where T4 primase can prime the synthesis of Okazaki pieces. Packaging of DNA is probably also initiated from recombinational or replicative forks (Mosig, et al., 1981) and requires concatemeric molecules created by recombination. All of the DNA in the T4 replicating pool may be part of a single intracellular concatemer of great complexity (Kemper and Brown, 1976).

1.2.c Gene expression

Transcription can be divided into two main phases, the pre-replicative period spanning the time from infection to the onset of DNA replication about 6 min later, and the post-replicative period which starts when DNA replication is initiated. Although host RNA polymerase is used throughout infection (Haselkorn et al., 1969; Goff and Weber, 1970), it is subject to chemical alterations and binds several T4 proteins. During the pre-replicative period host DNA, RNA and protein synthesis are shut off and early viral gene products appear. Those

gene products that appear for the first time after DNA replication starts are termed true-late, while those already being synthesised during the pre-replicative period but becoming more abundant afterwards are called quasi-late (O'Farrell and Gold, 1973).

(i) Pre-replicative transcription

T4 gp alt is injected with T4 DNA (Rohrer et al., 1975) and catalyses the transfer of ADP-ribose to one of the α subunits of the polymerase enzyme in a reversible process termed alteration which is complete 30 sec after the start of infection. A second event catalysed by the phage mod gene product starts 1-2 min into infection and involves the irreversible ADP-ribosylation of all α subunits (Goff, 1974). Neither process is required for T4 development, but alteration lowers the affinity of the polymerase for the σ subunit (Rabussay et al., 1972), and modification of the α subunits is directly responsible for a lowered rate of initiation at E.coli promoters (Mailhammer et al., 1975).

Immediately after T4 infection, transcription of a class of genes called immediate early (IE) is initiated at early promoters by unmodified polymerase or its altered form. This transcription does not require protein

synthesis. Transcription of a second class of early genes termed delayed early (DE) starts about 2 min after infection and does require protein synthesis (Young, 1975). Two different mechanisms regulate DE gene transcription. IE genes are proximal and DE genes distal to early promoters (Salser et al., 1970) and they are separated by potential Rho-dependent transcriptional termination sites (Goldberg, 1970; Jayaraman, 1972; Richardson, 1970). Readthrough of such terminators would allow expression of distal DE genes. This may be analogous to early transcription in λ where Rho mediates termination of RNA chains in the absence of λ gpN which allows RNA polymerase to transcribe past the termination sites. It is not known whether T4 codes for an N-like anti-termination protein. Secondly, some DE genes may be transcribed only on activation of a new class of promoters termed middle promoters (Mattson et al., 1974; Mattson et al., 1978). Middle transcription units therefore may encompass middle genes which are only accessible from middle promoters, and IE and DE genes which are also accessible from early promoters. Thus many pre-replicative genes are expressed in two modes, early from initiation at early promoters and later from middle promoters.

Typically these are genes whose products are required throughout infection (Brody, *et al.*, 1971). Pre-replicative transcription units are found mainly in two large blocks on the T4 map between 158-75 kb and 123-147 kb (Wood and Revel, 1976). Early genes code for functions involved in phage specific nucleotide and nucleic acid metabolism, binding to cell membranes, and alteration of host transcriptional and translational machinery. Pre-replicative transcription units are oriented in the same direction and almost all RNA synthesis in this period is 1-strand specific (or counterclockwise in relation to the T4 map) (O'Farrell *et al.*, 1980).

IE promoters are recognised by relatively unmodified *E.coli* RNA polymerase, and the sequences of some IE promoters (Table 1.1) show strong homology with -10 and -35 regions of *E.coli* promoters (Rosenberg and Court, 1979). Early transcription units produce truncated RNA molecules in the presence of chloramphenicol (Young, 1975). Therefore protein synthesis seems to be necessary for anti-termination at IE-DE gene junctions. In addition, synthesis of T4 gp *mot* is necessary for initiation of DE transcription (Mattson *et al.* 1974; Mattson *et al.* 1978). The -10 promoter regions of some genes expressed with DE characteristics are like those of *E.coli*

Table 1.1

Gene	Sequence in Region: -35	- 10
Early Promoters:		
30	TTTGACTGAGCT	TATAAT
ORF2 (next to gene 30)	TATTAAGCCCGG	TATAAT
ipIII	TACTTGAATAGA	TAAAAT
frd	TTGTGAAAAAGTCTG	TATTAT
Middle Promoters:		
1	AGAAGTTTAATGCTTC	TATAAT
rIIb	ATCAAATAATGCTTCA	TAAAAT
45	TTTAACGTTATTGCTT	TATAAT
32	CTCATATTGCTTA	TATTAT
43	TAAGCAAGGCTTCGGC	TATAAT
Ecoli promoters		
(consensus)	TTGACA	TATAAT

Table 1.1

Sequences of the non-transcribed strand of some presumed early and middle T4 promoters, from Brody et al. (1983). A line above the sequence indicates a possible homology to the E.coli consensus sequence. The conserved sequence seen in middle promoters is doubly underlined.

promoters (Rosenberg and Court, 1979). However the -35 regions show less homology to E.coli sequences and a new -35 region consensus sequence has been defined; $A^{T}_{A}TGCTT$, which starts between -35 and -30 (Brody et al., 1983) (Table 1.1). This new consensus may comprise the mot recognition site. In addition mot dependence of DE gene transcription is enhanced by modification of the α subunits of RNA polymerase early in infection which weakens the σ-core interaction (Khesin, et al., 1976; Schachner et al., 1971).

(ii) Post-replicative transcription

The post-replicative period starts 5-6 min into infection when T4 DNA replication is initiated. Transcription of true-late genes begins 1-2 min later (Young et al., 1980). By about 9 min into infection host transcription is no longer detectable (Kennell, 1970) and several minutes after this, true-late transcripts become the dominant species and a number of pre-replicative genes are shut off. True-late proteins include phage coat proteins and those required for phage assembly. Their mRNA is mainly transcribed from the r-strand (Guha et al. 1971).

Quasi-late genes continue to be transcribed into the post-replicative period and are usually those required throughout infection e.g. gp32 (Krisch et al., 1974). All quasi-late genes

probably have the same polarity as early genes i.e. are transcribed from the l-strand (Jayaraman and Goldberg, 1970). They are regulated in a non-uniform manner (Guha et al., 1971).

Transcription of true-late genes depends upon both modifications of the RNA polymerase and on a modified DNA template created during replication. Three T4 coded proteins are essential for late gene transcription in vivo; gp33, gp45 and gp55. These proteins copurify with the RNA polymerase but gp33 binds more loosely and it is possible to separate it by phosphocellulose chromatography (Horvitz, 1973). In vitro transcription systems have shown that T4 modified or host core polymerase transcribe late T4 genes in the absence of the σ factor (Rabussay and Geiduscheck, 1977b). Amber mutations in gp33 and gp55 generate transcription defects and it had been assumed that gp33 and gp55 together would be a σ factor substitute. However, Kassavetis and Geiduschek (1984) have found that T4 gp55 alone confers T4 late promoter specificity in the absence of σ. gp33 competes with the σ factor for the same binding site on the core polymerase and may assist gp55 by blocking the σ-core interaction. Two proteins of apparent M_r 15 kd and 10 kd (Stevens, 1972) also bind to late RNA polymerase. The 10 kd protein has an

anti-σ activity (Stevens, 1977) and interacts physically with σ and the core polymerase (Khesin et al., 1972). In addition the irreversible ADP-ribosylation of the polymerase α subunits has achieved a maximum about 5 min into infection and this again seems to lower the affinity between σ and the core polymerase (Khesin et al., 1976).

The DNA template for late gene expression must contain HMC, however a cytosine content of up to 20% does not affect true-late transcription (Kutter and Wiberg, 1969). The specificity for transcription of HMC-DNA is conferred by a T4 protein alc encoded by the T4 alc/unf gene (Snyder et al., 1976) whose mechanism of action is unknown. Multiple mutants unable to make HMC-DNA are blocked in late transcription. This block is relieved by a mutation in alc/unf. In addition, late transcription is coupled to DNA replication (Riva et al. 1970a). If replication is not allowed to start, late transcription is blocked, and blocking replication after it has started also suppresses late gene transcription (Lembach et al., 1969; Riva et al., 1970a). gp45 is essential for late gene transcription (Wu et al., 1975) and although it does not copurify with RNA polymerase, gp45 will bind specifically to immobilised modified core polymerase (Ratner, 1974). gp45 is also

part of the T4 replisome, suggesting that the T4 replication proteins affect late transcription directly through their interaction with gp45. It is possible to uncouple late transcription from replication in certain T4 multiple mutants altered in DNA polymerase, DNA ligase and gene 46 encoded exonuclease (Cascino et al., 1970; Riva et al., 1970b; Wu et al., 1975). Here the block in replication is accompanied by the introduction of breaks in the DNA. T4 late genes must somehow be made competent for transcription and this may occur during replication. It has been suggested that late promoters cannot be recognised in intact double-stranded DNA and have to be made accessible at replicative forks, perhaps through the interaction of gp45 (Ratner, 1974) or by melting of the DNA caused by nicks or gaps (Rabussay and Geiduschek, 1977b). Mutations which lead to replication-independent late transcription may be able to imitate such gaps.

Several T4 late promoters have now been mapped (Christensen and Young, 1982). An AT-rich upstream consensus (TATAAATA) spans the position of the E.coli -10 promoter consensus sequence, but they do not seem to share a -35 consensus sequence (Elliot and Geiduschek, 1984). This suggests that special initiation factors are

required for their recognition, however it is unknown how the RNA polymerase binding proteins are involved in promoter recognition except that gp55 confers late promoter specificity in the absence of σ, and various other factors reduce the affinity of σ for core polymerase. If binding of σ to RNA polymerase results in initiation at early or middle promoters, access of σ to the core may be controlled to prevent overproduction of early or middle RNA at the expense of late transcripts.

1.2.d Translation

T4 uses the host's translational machinery but again it employs a number of strategies to bias the system towards translation of T4 mRNA. T4 is able to block the translation of E.coli transcripts. The lac operon is transcribed during the first few minutes of infection but most of the transcript is not associated with ribosomes (Kennell, 1970). Furthermore, T4 infection prevents continued growth of pre-induced polysomes, indicating that T4 infection interferes with reassociation of ribosomes to host mRNA. Ribosomes from T4 infected cells are much less efficient at translating E.coli mRNA and bind much more poorly than those from uninfected cells (Hsu and Weiss, 1969; Dube and Rudland, 1970).

Not all mRNA species of an organism are translated at the same rate. The strength of ribosome

binding regions and accessibility of these regions are of major importance. In T4 infections, mRNA availability for translation is also regulated by the sequence of appearance of mRNAs. Initiation of translation in T4 infections utilises an AUG codon and a Shine-Dalgarno sequence (Gold et al., 1981). Mutations which narrow spacing between the Shine-Dalgarno sequence and the AUG codon on T4 rIIB mRNA severely reduce translation efficiency (Pribnow et al., 1981; Nelson et al. 1981), whereas increasing the spacing only has a small effect. Secondary or higher order structures can make the binding site more or less accessible to ribosomes. T4 rIIB is transcribed in two modes, early and middle, to give either polycistronic or monocistronic message (Daegelen et al. 1982) and the monocistronic form is more active in protein synthesis.

In addition, the T4 regA gene product is known to regulate the utilisation of many T4 early transcripts (Trimble and Maley, 1976). Reg A^- phage overproduce translational products of many early mRNAs e.g. gpe, and others are underproduced (Wiberg et al., 1973). A possible role for reg A protein is as a translational repressor which can recognise a structural feature common to all the mRNAs it affects. This feature may be a short sequence within the ribosome binding domain (Karam et al., 1981). Reg A seems to affect early mRNA

stability, which is greatly increased in reg A⁻ infections (Wiberg et al., 1973). This may however be an indirect consequence of inhibition of translation of reg A controlled mRNAs since the mRNA of T4 gene 1 is stable even though it is sensitive to inhibition by reg A (Trimble and Maley, 1976).

T4 codes for a number of components used in translation or which modify existing host translational components. None of these functions is essential for T4 development but some are lethal on certain bacterial strains (Guthrie and McClain, 1973). These include: a function that modifies existing tRNAs e.g. specific cleavage of host leucyl-tRNA (Yudelevich, 1971), functions that influence host tRNA modifying enzymes (Boezi et al., 1967), a function which alters valyl-tRNA synthetase (Muller and Marchin, 1975), functions that alter translation inhibitory factors (Rabussay and Geiduschek, 1977a) and also eight phage encoded tRNA species (McClain et al., 1972). These may help to bias the hosts translational machinery towards translation of T4 mRNA and T4 message codon usage.

1.2.e Particle assembly and cell lysis

T4 phage particle assembly is not programmed in time by successive expression of phage genes. All structural proteins seem to be synthesised simultaneously and accumulate as precursor pools from which are first constructed sub-assemblies and finally

complete virions. Overall assembly has been deduced by *in vivo* studies of mutant phage and by complementation in *in vitro* assembly (Edgar and Wood, 1966). Essentially capsid assembly consists of four major sub-assembly processes (Casjens and King, 1975). The base plate assembles first, and upon this assemble the tail tube and sheath. The shell of the phage head, composed mainly of gp23, is assembled separately. Once formed, the head and tail parts combine spontaneously. Formation of the tail fibres is independent of the rest of the virion, and they are attached to the base plate after the head and tail are joined.

DNA packaging occurs by a headful mechanism using concatemeric DNA as a substrate (Ritchie and White, 1972). Sequence specific cuts are not required as daughter molecules are circularly permuted and are 3% longer than the whole genome so that a complete genome is always packaged. A mutant with temperature-sensitive gp49 forms empty heads at non-permissive temperatures which become filled on lowering the temperature (Luftig *et al.*, 1971). Gene 49 is the structural gene for T4 endonuclease VII (Kemper *et al.*, 1981) which is able to remove recombinational (Holliday) structures from DNA (Mizuuchi *et al.*, 1982). In the absence of gp49 branched DNA accumulates (Kemper and Brown, 1976) which is partially packaged, presumably until

interrupted by a branch which would normally be removed by gp49. The mechanism of packaging termination is not understood.

A late gene product gpe is a lysozyme which cleaves bacterial cell wall peptidoglycans (Tsugita, 1971). Although lysozyme accumulates in the cell from just after the start of replication, it can only function if it has access to the bacterial cell wall. This is provided by the T4 gene t product (Josslin, 1970) which damages the cytoplasmic membrane. Cell lysis occurs usually about 30 min after the start of infection liberating a few hundred progeny (Ellis and Delbrück, 1939).

T4 exhibits lysis inhibition, where normal lysis is inhibited by secondary infection by T4 phage after the primary infection. The superinfecting phage are effective in lysis inhibition even if they have been killed by X-rays. The delay in lysis is a few hours long and about 1000 progeny are released. The cause seems to be a change in the cell membrane rendering it insensitive to gpt activity (Josslin, 1971). This phenomenon can often be used to increase the yield of phage in high titre lysates. Mutants in the rI, rII and rIII genes (Emrich, 1968) are not subject to lysis inhibition, and mutations in t resulting in lysis defective phage are suppressed by rII mutations. rII apparently prevents or delays gpt activity and hence delays lysis.

In rII⁻ phage, lysis is not subject to delay and occurs more rapidly than in wild type infections.

1.3 RESTRICTION ANALYSIS AND CLONING OF THE T4 GENOME

1.3.a. Preparation of cytosine-containing DNA

Wild-type T4 DNA cannot be cleaved by commonly used restriction endonucleases due to the presence of glucosylated HMC residues (Li et al., 1975; Kaplan and Nierlich, 1975). Presumably these residues protect target sites and prevent recognition, therefore DNA must be non-glucosylated in which case it is partially susceptible to some enzymes, or the HMC must be replaced by cytosine which is not a substrate for T4 glucosyl transferase (Kornberg et al., 1961). Three enzymes have recently been reported which cleave HMC-containing T4 DNA; TaqI, AhaIII and EcoRV, however their restriction patterns are different for T4 cytosine or HMC- containing DNA (Kutter and Rüger, 1983).

T4 HMC glucosylation is prevented by mutations in the structural genes for the α and β glucosyl transferases αgt and βgt (Georgopolous, 1967), or by using a host deficient in the UDP glucose phosphorylase, galU (Hattman and Fukasawa, 1963). Glucosylation protects T4 DNA from restriction by the rgl system of E.coli and the P1 phage system (Revel and Luria, 1970). Non-glucosylated T4 DNA is particularly susceptible to digestion by EcoRI (Kaplan and Nierlich, 1975).

T4 DNA containing cytosine can be made in several ways. A temperature-sensitive mutant of g56 (the dCTPase) forms DNA free of cytosine at 37°C, is lethal at 42°C, but at 39°C it produces phage with 20% of HMC replaced by cytosine (Kutter and Wiberg, 1969). Introduction of a denB mutation (endonuclease IV) prevents fragmentation of the cytosine containing DNA, and in order to make viable phage, strains must also have a mutation in the alc/unf gene. This allows transcription of late genes coding for capsid proteins from cytosine containing DNA. g56, denB, alc mutants can make viable phage in which cytosine replaces 40-70% of the HMC (Snyder et al. 1976). The cytosine content in the DNA can be raised to over 95% by including a denA (endonuclease II) mutation further reducing degradation of cytosine containing DNA (Warner, et al., 1970), or by introducing a gene 42 (dCMP hydroxymethylase) mutation (Morton et al., 1978; Wilson et al. 1977). T4 20% cytosine DNA is susceptible to partial cleavage by EcoRI but not HindIII, and susceptibility increases with the percentage of cytosine. DNA which is non-glucosylated or partially HMC-containing is useful for generating partial restriction digests (Wilson et al. 1977; Mattson et al. 1977). 100% cytosine-containing T4 DNA is susceptible to all restriction endonucleases tried so far.

1.3.b Mapping of the T4 genome

Most early restriction mapping of the T4 genome was achieved by ordering restriction sites contained on cloned T4 fragments carrying known genetic markers (e.g. Mattson et al., 1977; Velten et al., 1976; Wilson et al., 1977). These data together with analysis of digests of T4 cytosine DNA using a variety of multiple and partial digests (e.g. Takahashi et al., 1979; Kiko et al., 1979; Carlson and Nicolaisen, 1979; Kutter et al., 1980; O'Farrell et al., 1980; Rüger et al., 1979) have accumulated to provide a map that covers most of the genome. Figure 1.1 presents the genetic map of T4 correlated with the restriction map (Wood and Revel, 1976; Mosig, 1983). The construction of a purely genetic map of T4 is complicated by the extensive recombination undergone by T4 DNA molecules and local variations in recombination frequency (Stahl et al., 1964). Initially, the positions of many genes were determined by two-point crosses giving recombinational distances between conditionally lethal mutants (Edgar and Wood, 1966; Edgar et al., 1964; Stahl et al., 1964). The first estimates of physical distances between markers based on marker rescue with incomplete genomes (Mosig, 1968) showed that the recombinational map was distorted in some regions. Wood and Revel (1976) refined the relationship between the physical and

genetic distances by adjusting the map distances in several regions based on physical measurements of the sizes and locations of various deletions determined by heteroduplex mapping (Kim and Davidson, 1974) and the known molecular weights of gene products (O'Farrell et al., 1973; Vanderslice and Yegian, 1974).

The availability of a detailed restriction map is invaluable in the in vitro study of transcription of the T4 genome. Gram et al. (1984) used RNA polymerase from uninfected E.coli to transcribe normal T4 DNA and T4 cytosine-containing DNA. ^{14}C-labelled transcripts were separated by polyacrylamide gel electrophoresis. Transcripts were hybridised to restriction fragments and analysis of labelled bands allowed location of major transcripts larger than 0.5 kb. Transcripts from overlapping restriction fragments were analysed to give precise locations of promoters and Rho-dependent terminators relative to restriction sites, and to determine directions of transcription. A total of 31 strong promoter sites and 14 terminators were mapped. A number of T4 in vivo promoters have also been located by S1 nuclease mapping (e.g. Kassavetis and Geiduscheck, 1982). The pattern of transcription revealed that overlapping transcripts are observed throughout infection and that tandem promoters, first reported in the tRNA region by Goldfarb (1981a)

seem to be a common feature of the T4 early regions. This may improve the efficiency of T4 gene transcription relative to the host genome, or it may ensure that genes are transcribed in appropriate amounts during changes in the properties of the RNA polymerase (Goldfarb 1981b).

1.3.c The molecular cloning of genes in the pseT region

Cloning of sections of T4 DNA has already assisted in the study of gene expression and mapping of the genome, and has proved a valuable source of biochemically important enzymes e.g. DNA ligase (Wilson and Murray, 1979; Murray et al., 1979) and RNA ligase (Rand and Gait, 1984). Several groups have reported cloning of T4 DNA restriction fragments in λ and plasmid vectors, and have demonstrated the presence of numerous genes mainly by marker rescue analysis (Velten and Abelson 1980; Wilson et al., 1977; Vorozheikina et al., 1980). There is a bias against cloning of early genes which might be expected since early genes usually encode functions involved in DNA metabolism or host shut-off and would be detrimental to the host. In addition early genes are transcribed by relatively unmodified E.coli RNA polymerase, whereas expression of the late genes requires modification of the polymerase in order to recognise late promoters.

Wilson and Murray (1979) reported cloning of the intact functional DNA ligase gene (g30) in λ vectors using 50% cytosine DNA digested with EcoRI or HindIII. Fragments were isolated in integration proficient Red$^+$ λ replacement vectors or integration deficient Red$^-$ λ replacement vectors respectively. Recombinants including the T4 ligase gene were identified by their ability to complement the ligase deficient strain of E.coli, ligts7. At 37°C ligts7 produces insufficient ligase to allow Red$^-$ λ mutants to form plaques, and above 39°C the ligts7 bacteria fail to grow completely. Therefore λ T4 lig phage were detected as Red$^-$ λ able to form plaques on ligts7 at 37°C, or where the integration proficient vector was used, by the ability of a prophage to rescue host colonies at over 39°C. Lysogens of λ T4 lig phage in a suppressor-free host supported growth of T4 g30 amber mutants at high plating efficiency ($1-10^{-1}$), indicating that a functional g30 was expressed by the λ recombinant. Restriction analysis showed that g30 lies within a 1.9 kb HindIII fragment or 3 EcoRI fragments (0.4, 0.5 and 2.2 kb). While g30 is expressed only from λ promoters in recombinants carrying the 1.9 kb HindIII fragment, recombinants carrying all three EcoRI fragments include a T4 promoter able to initiate g30 expression. This promoter is present on the 1.9 kb fragment, but for unknown reasons it does not function. The g30 polypeptide

is transcribed anticlockwise in relation to the T4 map (Wood and Revel, 1976) as would be expected for a T4 early gene.

Murray et al. (1979) demonstrated that active ATP-dependent DNA ligase can be isolated following heat induction of E.coli cells lysogenic for λ T4 lig phage. This provides a genetic means of purifying ligase from other T4 proteins such as nucleases, and also results in amplification of expression.

Mileham et al. (1980) reported cloning and organisation of the frd-DNA ligase region of T4. Probes made from λ T4 lig and λ T4 td DNAs were hybridised to restriction fragments of 100% cytosine DNA. A single 11.5 kb HindIII fragment was identified which hybridised to both probes and therefore covered the region between them including the RNA ligase and polynucleotide kinase genes. The region from frd to g30 comprises 4 HindIII and 14 EcoRI fragments (Figure 1.5). Hybridisation of labelled fragments to partial EcoRI digests allowed unambiguous ordering of the EcoRI fragments. The 11.5 kb HindIII fragment was never cloned but six of the eight component EcoRI fragments and a deletion derivative (2.7 kb) of a seventh (3.1 kb) were recovered as λ recombinants. The EcoRI fragment (1.3 kb) and the deleted part of the 3.1 kb fragment which were not recovered may contain DNA encoding functions detrimental to the host or phage.

Figure 1.5

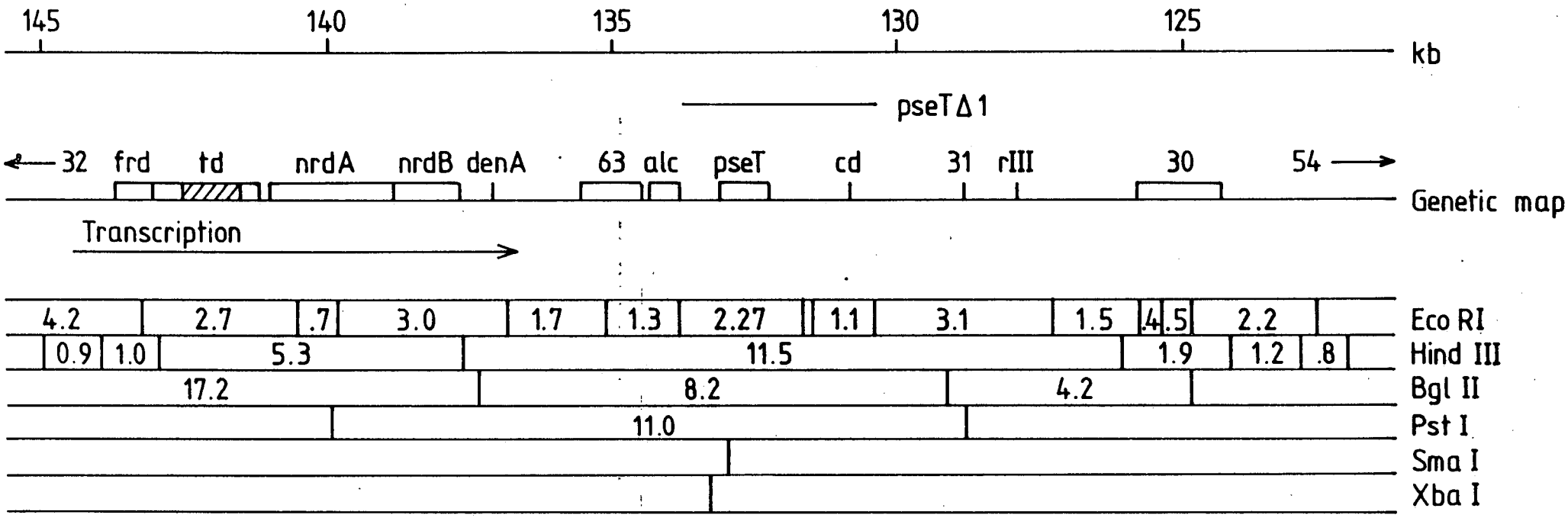

145
140
135
130
125
kb
pseTΔ1
32
frd
td
nrdA
nrdB
denA
63
alc
pseT
cd
31
rIII
30
54
Genetic map
Transcription
4.2
2.7
.7
3.0
1.7
1.3
2.27
1.1
3.1
1.5
.4
.5
2.2
Eco RI
0.9
1.0
5.3
11.5
1.9
1.2
.8
Hind III
17.2
8.2
4.2
Bgl II
11.0
Pst I
Sma I
Xba I

Figure 1.5

Organisation of the frd-DNA ligase region of the T4 genome. The figure shows: the physical distance in kilobase pairs from the rIIA-rIIB join = 0 (Wood and Revel, 1976); the map position of T4 deletion mutant pseTΔ1 as determined by heteroduplex mapping (Kutter et al. 1984); a genetic map of T4 showing the gene order in the frd-DNA ligase region, rectangles represent lengths of genes whose products have been identified and sized by SDS polyacrylamide gel electrophoresis, or by DNA sequencing; the direction of transcription in this region (Wilson and Murray, 1979; Mileham et al. 1980); a restriction map of the region showing fragment sizes (kb) for digestion with EcoRI, HindIII, BglII, PstI, SmaI, XbaI. The unlabelled EcoRI fragment has a size of 109 base pairs.

Other deleterious functions may require combinations of two or more fragments for their expression. Several genes are present in this region whose expression may be detrimental including denA, an endonuclease which attacks cytosine containing DNA, alc/unf which confers HMC-DNA specificity on transcription by the E.coli RNA polymerase, and the rIII gene whose product may be involved in a host membrane interaction. A functional rII gene cannot be cloned in plasmids (Selzer et al., 1978). In addition the presence of a strong promoter on a cloned fragment may interfere with transcription and replication of the vector.

The location of pseT is defined by the pseT$^-$ deletion mutant pseT Δ1 which deletes all known markers within this gene (Sirotkin et al., 1978). Mileham et al. (1980) using hybridisation of labelled EcoRI fragments to digests of T4 100% cytosine DNA, determined that the pseT Δ1 deletion spans 3 EcoRI fragments of 2.27, 0.1 and 1.1 kb (Figure 1.5). Either of the larger fragments could encode the polynucleotide kinase subunit (which has an apparent M_r of 33 kd), but if the gene extended over more than one EcoRI fragment, partial digests of T4 DNA should provide fragments including an intact pseT gene.

More recently Rand and Gait (1984) obtained a clone containing an intact functional RNA ligase gene. RNA ligase (gp63) (Snopek et al., 1977) again has many practical applications in experimental biochemistry

(Uhlenbeck and Gumport, 1982). Previous attempts to clone an intact gene 63 were unsuccessful and failure to obtain marker rescue of a g63 amber mutant with any of the λ recombinants from the frd-DNA ligase region had led Mileham et al. (1980) to conclude that the 1.3 kb EcoRI fragment might contain g63. An 8.2 kb BglII fragment (Figure 1.5) large enough to be sure of containing g63 was fragmented into 200-600 bp sections (Rand and Gait, 1984) which were cloned in M13 and the whole sequence of the fragment obtained. An open reading frame of the correct size occurred at the genetic map location of g63 (Wood and Revel, 1976; O'Farrell et al., 1980) and extended over the junction of the 1.3 and 1.7 kb EcoRI fragments with the 3' end in the 1.3 kb. One of the short fragments in M13 contained the 3' end of the open reading frame and this sequence was fused to a suitable restriction fragment, purified from T4 DNA, which contained the 5' region of the open reading frame. Most of the DNA surrounding the reconstituted g63 was removed by restriction before cloning g63 into mp8. The resulting clone expressed a functional RNA ligase. Rand and Gait (1984) suggest that a promoter is located in the 1.3 kb EcoRI fragment, 3' to g63, and a strong T4 promoter recognised by RNA polymerase in vitro maps in this region (Gram et al., 1984). Expression of the mutant T4 alc/unf gene downstream of this

promoter (Kutter et al., 1984) may explain why the 1.3 kb fragment could not be cloned intact.

1.4 EXPRESSION OF GENE PRODUCTS FROM λ RECOMBINANTS

1.4.a The λ genome

The temperate coliphage λ has about 50 genes, only half of which are essential for phage growth and plaque formation. Figure 1.6 shows the λ genetic map and transcriptional circuits (Friedman and Gottesman 1983; Arber, 1983). The left hand region includes all the genes (A to J) whose products are necessary for head and tail formation and DNA packaging. The region from J to the phage attachment site att contains non-essential genes coding for proteins of unknown function. Genes to the right of att govern site-specific (int and xis) and generalised (redA and redB) recombination of λ phage DNA. The product of the N gene is a regulatory protein necessary to activate transcription of λ genes including O, P and Q. None of the genes in the region between J and N is essential for plaque formation. Gene cI encodes the λ repressor protein which switches off transcription of phage genes in the lysogenic state (Ptashne, 1971). To the right of cI are genes which are essential for plaque formation. The O and P genes are required for replication of phage DNA, and the Q gene product for activation of late transcription. The S and R gene

Figure 1.6

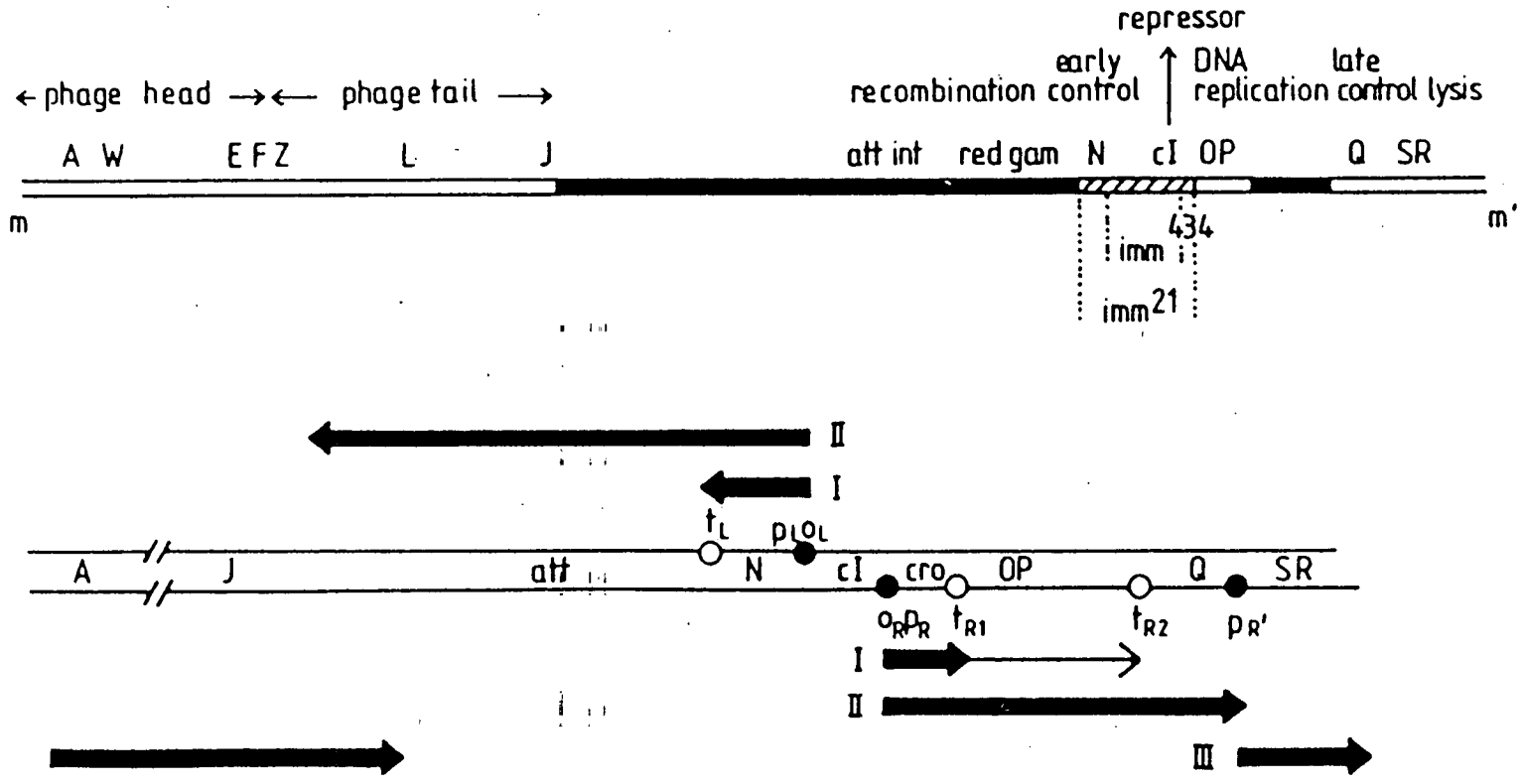

repressor
early
recombination control
DNA
replication control
late
lysis
← phage head →
← phage tail →
A W
E F Z
L
J
att int
red gam
N
cI
OP
Q
SR
m
m'
434
imm
imm21
II
I
tL
pLOL
A
J
att
N
cI
cro
OP
Q
SR
oRPR
tR1
tR2
pR'
I
II
III

Figure 1.6

A genetic map of λ including some key markers and transcriptional circuits, from Murray (1983). Black regions in the top diagram are non-essential, dotted lines indicate the end points of the immunity regions of the λ*imm*^{434} and λ*imm*^{21} phages, o_L and o_R are the *c*I repressor binding sites. The bottom diagram indicates organisation of the phage control region including sites at which transcription is initiated and terminated. Heavy arrows represent the major transcripts, filled circles are major promoters and open circles are major termination sites. (I) the immediate early transcripts initiated at p_L or p_R terminate at t_L and t_{R1} respectively in the absence of the *N* gene product. Some transcripts escape t_{R1} but terminate at t_{R2}. (II) in the presence of N protein, early transcripts initiated at p_L continue through t_L and those from p_R continue through t_{R1} and t_{R2}. (III) late transcription of the circular genome is dependent on the *Q* gene product and continues through genes *S R* and *A* and eventually *J*.

products are necessary for lysis of the host cell.

All the essential, and many of the non-essential genes are transcribed from the three major λ promoters p_L, p_R and p_R'. Following infection, transcription proceeds leftward from p_L through gene N and rightward from p_R through gene cro. In the absence of gene N product, most transcripts terminate just beyond these genes at sites t_L and t_{R1}. Those transcripts which escape termination at t_{R1} terminate at t_{R2}. The N gene product interacts with specific sites (nut) allowing RNA polymerase to overcome termination (Friedmann and Gottesman, 1983) and continue leftward beyond t_L into the non-essential genes and rightward through the DNA replication genes O and P and through t_{R2} into gene Q.

The product of gene cro is essential for lytic development and inhibits synthesis of the repressor (cI product). It also represses synthesis of early transcripts from p_L and p_R when the products of these genes are no longer required.

The product of Q like that of N is an anti-terminator, and is required for transcription of the late genes of λ. Since the linear λ chromosome circularises on infection, Q-dependent transcription from p_R, continues through S and R into A and on through J. The anti-termination activity of gpN is essential for activation of gene Q. However, if t_{R2} is deleted, leakage of transcription through

t_{R1} is enough to provide O, P and Q functions.

Depending on the conditions, 10-50% of infections with λ^+ phage give rise to lysogeny, where the repressed λ genome integrates into the chromosome of the host via a specific site on λ DNA called att. The prophage is trapped on the chromosome by the action of the λ cI repressor which binds to the p_L and p_R operator regions and prevents further transcription of the λ genes responsible for the lytic cycle (Wu et al., 1972). If the repressor is inactivated, the genome is expressed as described above and the prophage excises allowing replication and maturation of phage particles.

The large region of non-essential DNA in the genome (about 20 kb in λ wild-type) is central to the use of λ as a vector. In vector molecules at least part of the non-essential region is deleted to make space for the donor DNA. However, deletion of the entire non-essential region results in a loss of infectivity since encapsidation of λ DNA has a minimum as well as a maximum size limit (Weil et al., 1973). Usually molecules must be 80-105% of normal length, so infectivity can be maintained by replacing non-essential regions with other DNA sequences. A small fragment is more readily cloned in an insertion type λ vector. Such a vector has a single restriction target, preferably within a gene so that insertional

inactivation will allow recognition of recombinants by a changed but non-lethal phenotype (e.g. clear plaques are obtained on insertional inactivation of the cI gene since this prevents establishment of repression). DNA fragments greater than 10 kb can be cloned in a replacement vector, replacing non-essential DNA between widely spaced targets.

Packaging of linear λ DNA molecules can be achieved *in vitro* (Becker and Gold, 1975) using concentrated cell extracts of induced lysogens containing prophage with complementing capsid defects. *In vitro* packaging can be used under conditions which are size selective in order to enrich for recombinants generated by insertion of DNA into a vector with a small genome.

To achieve expression of cloned DNA, fragments are usually incorporated into the central non-essential region of λ between J and N so that transcription of the cloned genes is mediated by the major λ promoters p_L and $p_{R'}$. Genes inserted in the correct orientation downstream of p_L are transcribed from this promoter (Franklin, 1971). Inserted sequences do not usually impede transcription from p_L in the presence of the anti-termination activity of gene N (Adhya *et al.*, 1974; Franklin, 1974). Alternatively, genes inserted in the opposite orientation can be effectively transcribed from $p_{R'}$ (Wilson and Murray, 1979).

1.4.b. Amplification of gene products

A number of factors contribute to the efficiency of expression of cloned genes. Firstly, the coding sequence must be cloned without interruptions. This is not usually a problem for prokaryotic genes, however an intron has recently been discovered in the T4 td gene (Chu et al., 1984). For maximum amplification, transcription must be mediated by an efficient and tightly controlled promoter, particularly if accumulation of a particular gene product is likely to be detrimental to the host or vector (Hershfield et al., 1974; Shimitake and Rosenberg, 1981). Prokaryotic promoters share two regions of homology, one 35 base pairs upstream of the transcription initiation site (consensus 5'TTGACA) and another 10 base pairs upstream (consensus 5'TATAAT, the Pribnow box) (Rosenberg and Court, 1979; Siebenlist et al., 1980). These probably represent the bases most intimately involved in polymerase binding and orientation. The number of nucleotides separating the two consensus sequences is important. Altering the optimum distance of 16-19 nucleotides in the lac or β-lactamase promoters changes the 'strength' of the promoter. The most widely used systems for promoting high levels of cloned gene expression are: the lac promoter of E.coli which is regulated by the lac repressor and induced by IPTG; the trp promoter of E.coli which

is regulated by the trp repressor and induced by 3-indolylacetic acid or tryptophan starvation; the constitutive β-lactamase promoter of pBR322; and the leftward p_L and rightward p_R promoters of λ. The latter are regulated by the cI repressor, usually in a temperature sensitive cI mutant induced by a temperature increase.

Transcription termination is usually effected in a GC-rich region with 2-fold symmetry (a stem loop) before the termination site, followed by an AT-rich sequence at the actual site of termination (Rosenberg and Court, 1979). A number of protein termination factors are also involved, notably Rho (Roberts, 1970), as well as anti-termination factors such as gpN of λ (Greenblatt, 1981). Terminators may be inserted downstream of cloned genes to avoid adverse effects on replication or transcription of the vector (Nakamura and Inouye, 1982).

Amplification is usually augmented by increasing the gene copy number (Müller-Hill et al., 1968), either by cloning in a high copy number plasmid or in λ, where lytic infection allows multiplication of the recombinant phage. In the latter case, an integration-proficient λ phage can be propagated as a prophage where transcription from the major λ leftward and rightward promoters is repressed. Expression of a gene cloned in the centre of the λ prophage is initiated from p_L or p_R, only on induction of the

lysogen causing the prophage to excise and undergo multiple rounds of replication.

Efficient translation of transcripts depends upon the presence of a good ribosome binding site (rbs) which in E.coli usually consists of an initiation codon, AUG, and a sequence of 3-9 bases complementary to the 3' end of 16S rRNA (Shine and Dalgarno, 1975). This sequence is probably involved in binding of the 30S subunit to mRNA. Since an early T4 rbs is recognised by the E.coli translational machinery, translation of a cloned T4 early gene will presumably initiate efficiently from its own rbs.

The structural features of protein modification are not well understood. Bacterial proteins can undergo a number of modifications following translation e.g. hydrolysis of the N-terminal formyl methionine, or cleavage of hydrophobic signal sequences (Davis and Tai, 1980). It is possible that some phage encoded proteins may require modification by phage encoded functions. For example, glucosylation and phosphorylation do not occur to any great extent in E.coli.

The first systems employing λ to amplify expression of cloned genes involved insertion of a gene including its own promoter. The yield of polypeptide could be enhanced simply by delaying cell lysis so that DNA replication could increase the

number of gene copies, and the time available for expression would be extended. This was originally achieved by a mutation in the λ S gene (Müller-Hill et al., 1968) which prevents cell lysis but permits DNA replication and protein synthesis to continue for some hours. Mutations in Q and N proved even more efficient since these block all late functions including lysis and packaging of phage DNA (Moir and Brammar, 1976). This approach has been used to amplify both E.coli DNA ligase (Panasenko et al., 1977) and DNA polymerase I (Kelley et al., 1977). In both these cases the functional gene could not be cloned in a multicopy plasmid, presumably because overproduction of their protein products was deleterious.

Since cloned genes will not all have a functional promoter within the cloned sequence, Moir and Brammar (1976) investigated ways of optimising gene expression from p_L. Normally in λ, p_L mediates early leftward transcription (Figure 1.6) through gene N and beyond att. λ p_L is a very strong promoter, reportedly 8-10 times more efficient than lac (McKenney et al., 1981) in terms of transcriptional initiation. It is subject to control, primarily repression by the cI and the cro gene products. In addition, the nutL site downstream of the λp_L allows the N gene product in association with RNA polymerase to overcome transcriptional termination in sequences downstream of p_L (Franklin, 1974;

Adhya et al., 1974). Moir and Brammar (1976) used a cro⁻ λ in order to prevent modulation of transcription from p_L (Franklin, 1971; Sly et al., 1971) and also mutations in Q and S to inhibit lysis. Cells infected at high multiplicity with a λ trp cro⁻ Q⁻ S⁻ phage in which trp genes were transcribed from p_L contained up to 20% of their soluble protein as products of the trp operon. However cro⁻ phage are defective in DNA replication (Folkmanis et al., 1977), so a good amplification of gene copy number is not achieved. Consequently the use of a cro⁻ vector has not been very successful in some cases, e.g. induction of a cro⁻ λpolA prophage (Murray and Kelley, 1979) where the DNA polymerase gene was transcribed from p_L after inactivation of heat-labile repressor. In this case the DNA polymerase gene was transcribed more successfully from its own promoter in a cro⁺ phage where efficient DNA replication compensated for a lower rate of leftward transcription. A temperature-sensitive cro mutant is available (Matsubara, 1976), but it is difficult to optimise conditions for derepressed transcription from p_L and good amplification of copy number by DNA replication. Therefore, it is often easier to use p_L in a plasmid vector (see below).

Another alternative is to use the Q-dependent late promoter p_R, which is active after DNA replication.

An example is transcription of the T4 DNA ligase gene from p_R, in a phage defective in gene E, which encodes the major capsid protein, and in gene S to prevent lysis. However, transcription from p_R, must traverse a large part of the λ genome before reaching the inserted gene, resulting in over-expression of a large number of unwanted λ proteins.

λp_L can be cloned in a suitable plasmid vehicle in the presence of the repressor gene. Bacteria containing about 20 copies of such a plasmid can be grown in the absence of expression from p_L until subsequent inactivation of heat-labile λ repressor will allow transcription from p_L. One of the advantages of p_L in contrast to e.g. the *lac* promoter of *E.coli* is that sufficient λ repressor is produced from a single copy of the cI gene to repress transcription from multiple plasmid borne copies of p_L. A temperature-sensitive cI857 gene can be maintained either on the host chromosome or on a compatible plasmid so that transcription from p_L can be switched on by raising the temperature. Bernard *et al*. (1979) showed that heat induction of plasmids containing *trp*A downstream of p_L controlled by cI857, produced up to 6.6% of the soluble cell protein as the product of *trp*A.

Remaut *et al*. (1981) described plasmid expression vectors including λp_L together with the *nut*L sequence upstream of the insertion site for

foreign DNA. p_L was controlled by maintaining the plasmid in a defective, cro$^-$ λ lysogen containing a chromosomal cI857 gene and a functional N gene to provide anti-termination activity in conjunction with the nutL site. Synthesis of up to 10% of total soluble cell protein was achieved for trpA or β-lactamase, particularly where the plasmid was induced in an N-expressing host. Presumably anti-termination increases transcription efficiency. More recently Remaut et al. (1983) have constructed improved p_L vectors derived from runaway replication mutations (Uhlin et al., 1979) where heat induction inactivates cI857 repressor, but also leads to amplification of plasmid copy number. Here T4 DNA ligase expression exceeded 20% of the total soluble cell protein.

The use of plasmids carrying λp_L has been successful in amplifying expression of the λcII protein, a phage regulatory protein which is extremely toxic to E.coli (Shimitake and Rosenberg, 1981). A lysogen carrying a defective λ prophage expressing cI857 could be transformed at high efficiency with plasmids expressing the cII protein, whereas non-immune cells could not be transformed. Synthesis of cII protein as about 4% of total cell protein was achieved on induction by a temperature shift.

1.5 T4 POLYNUCLEOTIDE KINASE

Many T4 encoded enzymes have now been assigned physiological functions, but the *in vivo* role of polynucleotide kinase (pnk) has not been determined. However, the ability to catalyse the transfer of the terminal phosphate of ATP to a 5'-hydroxyl terminus of DNA or RNA has meant that the enzyme has found wide use in experimental molecular biology. Its main application has been in the investigation of DNA and RNA sequences as a means of labelling oligonucleotides with ^{32}P for their analysis by gel electrophoresis (Murray, 1973; Maxam and Gilbert, 1977). End-labelling of nucleotides has also proved invaluable in the synthesis of specific DNAs such as the gene for yeast alanine tRNA (Khorana *et al.*, 1972), which was only possible using labelled oligonucleotides to monitor ligation by T4 ligase.

Polynucleotide kinase was independently discovered in T2 (Novogrodsky and Hurwitz, 1966) and T4 (Richardson, 1965) as an activity which could catalyse the transfer of the γ-phosphate of ATP specifically to 5'-hydroxyl termini of DNA, RNA and oligonucleotides. No pnk activity has been found in uninfected bacteria, but similar activities are present in mammalian nuclei (Novogrodsky *et al.*, 1966). The T4 pnk also has an indigenous 3'-phosphatase activity (Cameron and Unhlenbeck, 1977; Sirotkin *et al.*, 1978). This activity was first discovered by Becker and Hurwitz (1967) but initially it was not realised that it was associated with the pnk enzyme.

The 3'-phosphatase activity strongly prefers DNA to RNA as substrate and selectively removes 3'-phosphoryl termini from DNA or deoxyribonucleotides (Becker and Hurwitz, 1967; Cameron and Uhlenbeck, 1977).

Depew and Cozzarelli (1974) screened isolates of heavily mutagenised T4 for those unable to induce the 3'-phosphatase activity. They found one such mutant, pseT1, which multiplied normally on all laboratory strains of E.coli suggesting that the activity was not essential. However, after testing a number of clinical strains of E.coli they found one, CT196, which was restrictive for pseT1. Unfortunately this host also had a reduced plating efficiency for wild-type T4, so CT196 was crossed with Hfr strains of E.coli K12 and a hybrid strain CTr5x was isolated which restricted T4 pseT1 but not T4 pseT^{+}. This allowed the isolation of another mutant, pseT2, and mapping of the pseT mutations between gene 63 and gene 31. Chan and Ebisuzaki (1970) had isolated a number of mutants unable to induce the 5'-kinase activity, but since they were unable to find a phenotype associated with the deficiency they could not map the mutations.

The first connection between the kinase and phosphatase activities was made by Cameron and Uhlenbeck (1977) who reported that a 3'-phosphatase activity co-purified with 5'-pnk, and that both seemed to be associated with the same polypeptide. Sirotkin et al., (1978) showed that most pseT mutants, isolated because they were

restricted on CTr5x, lacked both the 5'-kinase and 3'-phosphatase activities. In fact Depew and Cozzarelli were unfortunate in not discovering that pseT is the gene for pnk since they assayed cells infected with their pseT1 mutant for kinase activity and found it to be normal. So far, pseT1 is the only mutation isolated which inactivates the 3'-phosphatase but not the 5'-kinase. A second mutant pseT47 fails to induce 5'-kinase and not the 3'-phosphatase. These two mutants both fail to make late proteins in CTr5x and do not complement each other for this defect, suggesting that the two functions are required for normal late expression in this strain (Sirotkin et al., 1978).

Detailed studies have been made of the polynucleotide kinase activity (Kleppe and Lillehaug, 1979; Richardson, 1965) and the physicochemical properties of the protein (Lillehaug, 1977; Panet et al., 1973). SDS gel electrophoresis of the homogeneous protein gives one polypeptide band of an apparent M_r of 33 kd (Panet et al., 1973; Lillehaug, 1977). The molecular weights of the active enzyme and the subunit have also been estimated from sedimentation equilibrium data as 147 kd and 33.2 kd respectively (Lillehaug, 1977). It seems therefore that the active species consists of 4 subunits. N-terminal amino acid analysis (Lillehaug, 1977) showing that phenylalanine is the only N-terminal residue, suggests that the subunits are identical. Analytical ultracentrifugation experiments show that in low ionic strength

buffer the enzyme dissociates into 4 subunits (Lillehaug, 1977), while high ionic strength or the presence of polyamines or the enzyme substrates favours stabilisation of the tetrameric form and stimulates kinase activity (Lillehaug and Kleppe 1975a and 1975b).

The kinase activity is almost lost in the absence of sulphydryl agents such as 2-mercaptoethanol, indicating that the enzyme is dependent upon reduced -SH groups for activity. Each monomer appears to have two -SH groups, one exposed on the surface and one more buried. The enzyme precipitates in the absence of 2-mercaptoethanol at low ionic strength suggesting that -SH groups are involved in stabilisation of the tertiary structure (Lillehaug, 1977). However, the determination of the number of -SH groups was not affected by the presence of the substrates (ATP or 3'-monophosphate), so cysteine residues are probably not part of the active site. The enzyme also requires Mg^{2+} for activity. The circular dichroism spectrum of the protein indicates a highly helical organisation. The α-helical content was estimated to be between 45 and 55% and β-pleated sheet about 25% (Lillehaug, 1977).

Substrates of the kinase reaction e.g. $(Up)_4U$, protect both the 3'-phosphatase and the 5'-kinase activities from heat inactivation, while compounds which are only substrates for the 3'-phosphatase e.g. $p(Up)_5$ cannot protect the 5'-kinase activity (Cameron and Uhlenbeck, 1977). This indicates that the activities

may be closely connected on the protein but are not manifestations of the same reaction. The existence of the pseT1 and pseT47 mutants which lack the 3'-phosphatase and the 5'-kinase activity respectively (Cameron et al., 1978; Sirotkin et al., 1978) again suggests that separate active sites exist.

The 5'-polynucleotide kinase activity can phosphorylate DNA, RNA and oligonucleotides. While mononucleosides are not substrates, a nucleoside 3'-monophosphate will be phosphorylated to yield the 3'-5'-diphosphate (Lillehaug and Kleppe, 1975a). Sano (1976) reported that CTP and GTP are equally effective as phosphate donors, as is ATP. The kinase prefers single-stranded 5'-hydroxyl termini, so blunt-ended or 5'-termini with 3' overhangs, or single-stranded nicks in DNA are poor substrates (Lillehaug et al., 1976). The 5'-kinase reaction is reversible (Van de Sande et al., 1973) [γ-^{32}P]ATP is produced from a 5'-[^{32}P] labelled oligonucleotide in the presence of ADP. However, the forward reaction has maximal activity at pH 9-10 while the reverse reaction has a pH optimum of 6.

The 3'-phosphatase activity selectively removes 3'-phosphoryl groups from DNA or deoxyribonucleotides (Depew and Cozzarelli, 1974) and will use a mononucleoside 3'-phosphate as a substrate yielding the nucleoside (Becker and Hurwitz, 1967; Cameron and Uhlenbeck, 1977). The activity is also a cyclic 2'-3'-phosphatase, however, there is little effect on 3'-phosphoryl RNA.

The optimum pH for the 3'-phosphatase reaction is pH 6 (Becker and Hurwitz, 1967).

The dual activities of the enzyme suggest that both the 5'-kinase and the 3'-phosphatase are involved in the same pathway *in vivo*. They have the potential to convert 3'-phosphate or 2',3'-cyclic phosphate, and 5'-hydroxyl terminated polynucleotides into substrates for RNA ligase, which requires 5'-phosphate and 3'-hydroxyl ends. There is some similarity between this model and the eukaryotic RNA splicing enzymes where RNA is cut next to an intron to yield a cyclic 2',3'-phosphate. A phosphatase cleaves this cyclic phosphate to a 2'-phosphate, and the 5'-hydroxyl end is phosphorylated to yield a 5'-phosphate terminus (Knapp *et al.*, 1977). The termini are then ligated to give a polynucleotide with a 2'-phosphate branch (Konarska *et al.*, 1982). The T4 phage encodes enzymes with all these activities except an intron-specific endonuclease to make the first cut. Also the phage encoded phosphatase cuts both the 2' and 3' linkages of the cyclic phosphate so that the final polynucleotide would not have a 2'-phosphate branch (Uhlenbeck, cited in Snyder, 1983). T4 *pseT* and the RNA ligase gene (g63) are closely linked on the genetic map (Figure 1.1) as are many T4 genes of related function.

E.coli CTr5x, the restrictive host for *pseT* is also restrictive for RNA ligase mutants (*rli*) (Runnels *et al.*, 1982) and the defects exhibited by *pseT* and *rli*

mutants are similar (see below). pseT and rli mutants also share a common suppressor stp (Depew and Cozzarelli, 1974; Runnels et al., 1982), a T4 gene which maps close to rII. A mutation in stp relieves the effects of rli and pseT mutants. Depew and Cozzarelli (1974), found that pseT mutations reduced the rate of T4 DNA replication by half in E.coli CTr5x at 37°C, and that the T4 DNA made was shorter than normal and had an increased frequency of interruptions. They postulated that 3'-phosphatase groups introduced into T4 DNA, possibly by an endonuclease coded for by the stp gene could be repaired by the host repair system in a normal E.coli strain, but not in CTr5x. Therefore in CTr5x the phage encoded 3'-phosphatase would be essential for removing 3'-phosphate ends which are not substrates for DNA ligase, polymerases or many exonucleases and are known to inhibit several of these activities (e.g. Goulian et al., 1968; Lehman and Nussbaum, 1968). However it has not been proven that 3'-phosphoryl termini occur in T4 infected cells in vivo, and no mechanism is known which produces 3'-phosphoryl termini in uninfected cells.

Sirotkin et al. (1978) and Runnels et al. (1982) confirmed the observation of defects in T4 DNA replication in pseT or rli mutants, but they also noticed a defect in late gene expression which was 3-5 fold lower than normal even though early protein synthesis remained unchanged. The defect in T4 DNA replication occurs before there is any effect on gene expression suggesting

that pnk and RNA ligase affect replication directly, rather than the appearance of an early gene product involved in replication. However, such a protein may have escaped detection on gels (Runnels et al., 1982). The effect on late gene expression seems to be at least partly at the transcriptional level (Sirotkin et al., 1978). Mixed competitor hybridisation studies with RNAs from T4 pseT2 infected CTr5x showed that RNA late in infection was deficient in sequences equivalent to normal T4 late mRNA. Usually when T4 replication is blocked by a mutation in a phage gene, the true-late gene products fail to appear but synthesis of early gene products continues. However, late in infection of CTr5x with pseT or rli mutants, synthesis even of early proteins stops so that little or no protein synthesis can be detected in the late period (Runnels et al., 1982). This is similar to the situation observed when T4 DNA containing cytosine replicates in the presence of an active alc/unf gene (Snyder et al., 1976), or when wild type T4 infects a host with a mutation in the lit gene (Cooley et al., 1979). E.coli lit mutants are defective in their ability to support late (replication-coupled) gene expression of T4 at 30°C, and some lit mutants restrict the growth of pseT mutants, but not wild-type T4, at 37°C. However in the case of cytosine-containing T4 mutants or host lit mutations, only T4 late gene expression and not DNA replication are affected.
In all these situations the effect on late gene expression

is more severe at low temperatures. The defects in late gene expression and DNA replication which arise on infection of CTr5x with pseT or rli mutants only occur when cells are grown and infected below 37°C. At 30°C, replication is about 10% of normal while there is almost no synthesis of late proteins. At 42°C gene expression and phage production are normal in CTr5x. Therefore, several lines of evidence suggest that the closely linked pseT, g63 and alc/unf genes of T4, and the host lit gene, may all affect the same requirement for T4 gene expression. Possibly they participate in the accumulation during replication of a DNA template which is competent for late gene expression.

A different theory of pnk and RNA ligase activity in vivo has also gathered considerable popularity. David et al. (1979), labelled permeablised T4-infected E.coli cells with [γ-^{32}P] ATP and found labelling of RNA which they presumed was mediated by the T4-induced pnk enzyme. The RNA was still labelled after treatment with alkaline phosphatase to remove end phosphates suggesting that ^{32}P was 'trapped' in the RNA after ligation of ends by the T4-induced RNA ligase. The label was located specifically in small tRNA-sized RNAs presumably of host origin. The infection of E.coli CTr5x resulted in the isolation of some cleavage products of a tRNA-like RNA (David et al. 1982). Four oligonucleotides were obtained, at least two of which could be put together to form a tRNA-like molecule with an anticodon

for isoleucine (David et al., 1982). The cleavage site was next to the anticodon in a position equivalent to cleavage during removal of introns in some species of yeast tRNA (Goodman et al., 1977). The cleavage fragments persist later into infection of CTr5x with pseT or rli mutants than in $T4^+$ infections, and they do not appear on infection with stp^- mutants. Again this suggests that stp encodes a nuclease which in this case cleaves the isoleucine tRNA of E.coli CTr5x, thereby blocking translation in general unless religation is affected by the action of T4 pnk and RNA ligase activities. However this does not explain why defects in infections with pseT or rli mutants are more pronounced at low temperatures.

More recently Chu et al. (1984), and Belfort et al. (1985) have described a T4 gene which produces a transcript including an intron within the coding sequence. This gene, td, encodes thymidilate synthase which is part of a multi-enzyme complex converting NDPs and dNDPs, released from degradation of the host genome, to dNTPs. Nucleotide sequence analysis of the td gene revealed an intron of 1017 bases interrupting the coding sequence for the mature protein (Chu et al., 1984). The mechanism by which processing to remove the intron occurs is not yet understood. However the cloned td gene is expressed in E.coli in the absence of T4 and the mature protein is produced (Chu et al. 1985). Free intron RNA appears on T4 infection even in the presence of chloramphenicol, however the production of mature td transcripts is inhibited

by chloramphenicol (Belfort et al., 1985). This indicates that intron splicing is independent of protein synthesis while efficient exon ligation may be dependent on host or phage protein synthesis. However, if pnk and RNA ligase are involved in processing, there must be an alternative mechanism in uninfected E.coli. So far only a 2'-5' E.coli RNA ligase has been reported (Greer et al., 1983) and a 2',5' linkage seems unlikely since sequencing with reverse transcriptase proceeds normally across the td mRNA splice junction (Belfort et al., 1985). The E.coli CTr5x strain may have a mutation in a host RNA ligase. Alternatively it may encode the tRNA-like substrate of RNA ligase described by David et al. (1982).

One possibility which may connect the apparently separate affects on DNA metabolism and RNA processing is that it is not the pnk and RNA ligase which act upon DNA, but rather an RNA which they process. The enzymes may be involved in processing several RNAs, only some of which are essential for T4 development.

1.6 AIMS AND STRATEGIES

The aim of this work is to identify the DNA sequence encoding the T4 pnk gene (pseT) and to clone the intact coding sequence with a view to amplifying expression of the protein product.

The pseT gene has been mapped between g63 (RNA ligase) and g31 (Depew and Cozzarelli, 1974). The T4 mutant pseTΔ1 which has a deletion of 3.5 kb (Kutter et al., 1984) has lost all known markers of pseT and lacks

three EcoRI fragments of 2.27 kb, 0.1 kb and 1.1 kb in this region (Mileham et al., 1980). These three fragments were cloned in λ vectors by Mileham et al. (1980), and the recombinants provide a source of DNA for the initial nucleotide sequencing of the pseT region.

The recognition of recombinants including pseT cannot rely on marker rescue techniques since the pseT gene is non-essential to T4. It is therefore necessary to rely on cloning a functional gene and to screen for kinase activity in infected cells. Cloning of the pseT gene in either a λ or plasmid vector would separate the kinase gene from T4 genes that encode nucleases, and further manipulation should facilitate high levels of expression of the enzyme.

The availability of quantities of homogeneous protein would assist in analysis of the structure and mechanisms of the enzyme by both physical and biochemical techniques such as X-ray crystallography and substrate binding kinetics. Additionally cloned genes are more readily subjected to specific site-directed mutagenesis, and analysis of altered proteins may indicate which sequences or structures are necessary for the 5'-kinase and 3'-phosphatase activities.

C H A P T E R 2

MATERIALS AND METHODS

2.1. MATERIALS

(a) Media

L-broth

Difco Bacto Tryptone	10 g
Difco Bacto yeast extract	5 g
NaCl	10 g
Distilled water	to 1 litre

This was adjusted to pH 7.2 with NaOH before autoclaving.

L-agar

Difco Bacto Tryptone	10 g
Difco Bacto yeast extract	5 g
NaCl	10 g
Difco agar	15 g
Distilled water	to 1 litre

This was adjusted to pH 7.2 with NaOH before autoclaving.

BBL agar

Baltimore Biological Labs. trypticase	10 g
NaCl	5 g
Difco agar	10 g
Distilled water	to 1 litre

BBL top agar

As for BBL agar but only 6.5 g Difco agar per litre.

Phage Buffer

KH_2PO_4	3 g
Na_2HPO_4 (anhydrous)	7 g
NaCl	5 g
0.1 M $MgSO_4$	10 ml
0.01 M $CaCl_2$	10 ml
1% w/v Gelatin	1 ml
Distilled water	to 1 litre

5 x Spizizen salts

$(NH_4)_2SO_4$	10 g
K_2HPO_4	70 g
KH_2PO_4	30 g
$Na_3C_6H_5O_7.2H_2O$ (tri-sodium citrate)	5 g
$MgSO_4.7H_2O$	1 g
Distilled water	to 1 litre

Minimal agar

Difco agar	6 g
Distilled water	to 300 ml

80 ml sterile 5 x Spizizen salts were added after autoclaving and the agar supplemented with 0.2% glucose and 2 µg/ml vitamin B_1 and made up to 400 ml with sterile water.

4 x M9 Salts

Na_2HPO_4	28 g
KH_2PO_4	12 g
NaCl	2 g
NH_4Cl	4 g
Distilled water	to 1 litre

M9-Maltose Medium

4 x M9 salts	250 ml
20% maltose	15 ml
1 M $MgSO_4$	1 ml
Distilled water	to 1 litre

All media were sterilised by autoclaving at 15 lb/square inch for 15 min.

2.1.b Enzymes and Chemicals

DNA polymerase (Klenow fragment) was purchased from Boehringer, T4 DNA ligase from New England Biolabs., DNA polymerase I from NBL Enzymes, restriction endonucleases from New England Biolabs., or Bethesda Research Laboratories, RNAses I and T_1, calf intestinal phosphatase, lysozyme and micrococcal nuclease from Sigma, spleen phosphodiesterase, pancreatic DNAse and venom phosphodiesterase from Worthington.

M13 17-mer primer and hybridisation probe primer were from New England Biolabs., cytidylyl-(3'-5')-uridine (CpU) from Sigma, deoxynucleoside triphosphates (dNTPs) and dideoxynucleoside triphosphates (ddNTPs) from P-L Biochemicals. Synthetic oligonucleotides were provided by E. Kawashima (Biogen, Geneva) and J. Keyte (Biochemistry, Leicester).

L -[^{35}S]methionine (specific activity > 30 TBq/nmol), Deoxycytidine 5'-[α-^{32}P]triphosphate (~110 TBq/nmol), Adenosine 5'-[γ-^{32}P]triphosphate (~110 TBq/nmol) were from Amersham International. Deoxyadenosine 5'-[α^{35}S]-thiotriphosphate (18.5 TBq/nmol) was from New England Nuclear.

DEAE-sephadex (A50 and G50) was from Pharmacia, DEAE-cellulose (DE52), phosphocellulose (P11) and AE-cellulose paper (AE81) from Whatman. 'Dowex' 50W-x8 (H) was from BDH. Nitrocellulose filters were from Schleicher and Schüll, HP5 film from Ilford, Cronex intensifier screens and X-ray film from Du Pont. All chemicals were Analar grade.

2.1.c. Bacterial Strains

E.coli K12	Relevant features and use	Reference
CR63	*supD*; permissive host for T4 amber mutants	Appleyard *et al*. (1956)
C600	*supE*; amplification of λ *Wam Eam Sam* phages	Appleyard (1954)
NM594	*sup*°; non-permissive host for λ amber mutants	Weigle (1966)
ED8689	*sup*°*hsdR*; non-permissive host for amber mutants	Wilson *et al*. (1977)
ED8654	*supE supF hsdR*; permissive host for λ amber mutants	Murray *et al*. (1977)
NM538	*supF hsdR*; permissive host for λ amber mutants	Frischauf *et al*. (1983)
NM514	*hfl hsdR*; selection of cI^- recombinants	Murray (1983)
NM522	(*lac-pro*)Δ*hsdMS*Δ F'*lacZM15lacIq*; host for M13 phages	Gough & Murray (1983)
M159	*sup*° *uvrA*; host for analysis of polypeptides	Jaskunas *et al*. (1975
C600 (λCM21)	*supE* (λCM21); source of λCM21 DNA	
ED8689 (λCM21)	*sup*° *hsdR* (λCM21): amplification of pnk	
C600 (λ)	*supE* (λ); selection of λimm^{434} phage	
C600 (λimm^{434})	*supE* (λimm^{434}); selection of λimm^{λ} phage	

2.1.d. Phage Strains

(i)

λ vectors and derivatives	Relevant features and/or use	Reference
NM607	Immunity insertion vector for EcoRI fragments	Murray et al. (1977)
NM1207	NM607 including T4 fragment 46	Mileham et al. (1980)
NM1210	NM607 including T4 fragment 21	" " " "
NM459	(srIλ1-2)Δ cI857 nin5 vector for EcoRI fragments	
NM1104	NM459 including T4 fragment 21	
CM1	NM459 with fragment 21 in opposite orientation to NM1104	
CM6	Deletion derivative of NM1104	
NM1108	Wam403 Eam1100 Sam100 derivative of NM1104	
NM1149	Immunity insertion vector for EcoRI or HindIII fragments	Murray (1983)
CM8	NM1149 including fragment 21+	
NM1070	Wam403 Eam1100 lacZ cI857 nin5 Sam100	
CM21	NM1070 including fragment 21+ in place of lacZ	
NM1112	b519 Qam73 Sam7; SmaI left arm for CM6	
NM1222	Eam1100 NM459; EcoRI right arm for CM6	

+ Alternative clone of fragment 21

(ii) λ phage	Relevant features and/or use	Reference
λ clear	cI26; selection of λ lysogens	
λ vir	virulent derivative of λ; test for λ sensitivity	
λ^+	λ immunity	
λ imm^{434}	imm^{434}	

(iii) T4 phage		
T4D	$T4^+$	H. R. Revel (pers. commun.)
pseT2	Defective in polynucleotide kinase polypeptides	Sirotkin, et al. (1978)
am N122	$g42^-$; $pseT^+$	Hughes & Brown (1973)
alc7	alc7, am51 (g56), NB5060 (denB-rII), am87 (g42); 100% cytosine DNA	Wilson et al. (1977)

(iv) M13 vectors

M13mp8, M13mp11 (Messing, 1983) and M13mp18 (Yanisch-Perron et al., 1985) were used as vectors for DNA sequencing.

2.1.e. Solutions

TE buffer:

10 mM Tris, 1 mM EDTA adjusted to appropriate pH with HCl.

Hanahan competent cells:

TFB buffer

10 mM K-MES pH 6.2

100 mM RbCl or KCl

45 mM $MnCl_2.4H_2O$

10 mM $CaCl_2.2H_2O$

3 mM $HACoCl_3$ (Hexamino Cobaltic chloride)

1 M MES was adjusted to pH 6.3 using KOH, sterile filtered and stored at -20°C. TFB was filtered and stored at 4°C.

DMSO

Spectroscopically pure dimethyl sulphoxide was saturated with nitrogen and stored at -70°C in small aliquots.

DTT

2.25 M DTT in 40 mM Potassium acetate, pH 6.0 was filtered and stored at -20°C.

Packaging reaction:

Buffer A

20 mM Tris.HCl pH 8.0

3 mM $MgCl_2$

0.05% (v/v) 2-mercaptoethanol

1 mM EDTA pH 7.5

Buffer M1

110 µl H_2O

6 µl 0.5 M Tris.HCl pH 7.5

300 µl 0.05 M Spermidine, 0.1 M putrescine, neutralised with Tris base

9 µl 1M $MgCl_2$

75 µl 0.1 M ATP

1 µl 2-mercaptoethanol

Filter hybridisation:

Denaturation buffer

0.5 M NaOH

1.5 M NaCl

Neutralisation buffer

0.5 M Tris

3.0 M NaCl

neutralised to pH 7.0 with 42ml/litre HCl

20 x SSC

3.0 M NaCl

0.3 M $Na_3C_6H_5O_7$

1 x dNTP buffer

4 x buffer	210 mM Tris.HCl pH 7.5
	21 mM $MgCl_2$
	20 µg/ml BSA
1 x buffer	100 µl 4 x buffer
	4 µl 2mM dATP/dTTP/dGTP
	1 µl 2-mercaptoethanol
	295 µl Distilled water

Hybridisation buffer

50% Formamide

4 x SSC

1 x Denhardt solution

20 x Denhardt solution

0.4% w/v BSA

0.4% w/v Ficoll

0.4% w/v Polyvinylpyrrolidone

Gels for DNA:

10 x TBE Buffer

432 g Tris

220 g Boric acid

37.2 g EDTA

made up to 4 litres with distilled water

TBE loading dye

20 g Sucrose

0.1 g Bromophenol blue

10 ml 10 x TBE

made up to 100 ml with distilled water

30% Acrylamide stock

29 g Acrylamide

1 g N,N'-methylene bisacrylamide

made up to 100 ml with distilled water

Polyacrylamide gels

	%	3.5	5.0	12.0	20.0
30% acrylamide	(ml)	11.6	16.6	40.0	66.6
distilled water		73.6	71.3	47.9	21.3
3% ammonium persulphate (AMPS)		2.1	2.1	2.1	2.1
10 x TBE		10.0	10.0	10.0	10.0
Range of separation (nucleotides)		100-1000	80-500	40-200	10-100

DNA sequencing:

TE buffer

10 mM Tris

0.1 mM EDTA

adjusted to pH 8.0-8.5 with HCl

TM buffer

100 mM Tris

50 mM $MgCl_2$

adjusted to pH 8.5 with HCl

Termination mixes

Stock 10 mM:	ddTTP	61 mg/10 ml
	ddCTP	58 mg/10 ml
	ddGTP	62 mg/10 ml
	ddATP	62 mg/10 ml
Stock 50 mM:	dTTP	312 mg/10 ml
	dCTP	296 mg/10 ml
	dGTP	316 mg/10 ml
	dATP	295 mg/10 ml

all made up in TE buffer.

Termination mix

	T	C	G	A
0.5 mM dTTP (µl)	25	500	500	500
0.5 mM dCTP	500	25	500	500
0.5 mM dGTP	500	500	25	500
10 mM ddTTP	50			
10 mM ddCTP		8		
10 mM ddGTP			16	
10 mM ddATP				1
TE	1000	1000	1000	500

Sequence chase mix

0.25 mM dTTP

0.25 mM dCTP

0.25 mM dGTP

0.25 mM dATP

made up in TE buffer using the 50 mM stocks

Klenow polymerase mix

per clone: 4 µCi [γ-^{35}S]ATP,
1.5 units Klenow polymerase
in a total volume of 9 µl with
10 mM Tris.HCl pH 8.5, 10 mM
DTT. The appropriate quantity
of mix was made up immediately
before dispensing into reactions.

Formamide dyes

100 ml De-ionised formamide

0.1 g Xylene cyanol FF

0.1 g Bromophenol blue

2 ml 0.5 M EDTA

40% Acrylamide stock

380 g Acrylamide

20 g N,N'-methylene bisacrylamide

made up to 1 litre with distilled water and de-ionised.

0.5 x TBE gel mix

150 ml 40% Acrylamide

50 ml 10 x TBE

460 g Urea

made up to 1 litre with distilled water

2.5 x TBE gel mix

150 ml 40% Acrylamide

250 ml 10 x TBE

460 g Urea

50 g Sucrose

50 mg Bromophenol blue

made up to 1 litre with distilled water.

SDS-PAGE for proteins:

SDS sample buffer

2 ml Glycerol

0.4 g SDS

5 ml 0.25 M Tris.HCl pH 6.8

2 ml Distilled water

1 ml 2-mercaptoethanol (added just before use)

Acrylamide Stock solution

44 g Acrylamide

0.3 g N,N'-methylene bisacrylamide

Separating gel buffer (A)

0.75 M Tris.HCl pH 8.8

0.2% w/v SDS

Spacer gel buffer (B)

0.25 M Tris.HCl pH 6.8

0.2 % w/v SDS

Separating gel mix

13.5 ml Gel buffer A

9.2 ml Acrylamide stock

3.6 ml Distilled water

0.95 ml 10 mg/ml AMPS

75 µl TEMED

Stacking gel mix

10 ml	Gel buffer B
3.3 ml	Acrylamide stock
6.7 ml	Distilled water
0.5 ml	10 mg/ml AMPS
40 µl	TEMED

10 x Laemmli buffer

30 g	Tris
144 g	Glycine
10 g	SDS

made up to 1 litre with distilled water.

Lowry-Folin assay:

Reagent A: 2% Na_2CO_3 (anhydrous) in 0.1N NaOH

" B: 0.5% $CuSO_4.5H_2O$ in 1% $Na_3C_6H_5O_7$

" C: 1 ml of B + 50 ml of A

" D: Folin-Ciocalteu Reagent diluted 1:1 with water.

2.2 METHODS

2.2.a. Plating cells

A fresh overnight culture was diluted 20 fold into L-broth and grown at the required temperature for 2 h. Cells were pelleted by spinning for 10 min at 3,000 rpm in a bench centrifuge and resuspended in a half volume of 10 mM $MgSO_4$ before storage at 4°C.

2.2.b. λ and T4 phage plate lysates

A single plaque was picked into 1 ml of phage buffer containing a drop of chloroform, mixed and left for 10 min. 0.1 ml of the suspension was added to 0.1 ml of a fresh overnight supplemented with 10 mM $MgSO_4$ or 0.2 ml of plating cells and the phage adsorbed for 10 min before plating out on to a fresh wet L-agar plate in 2.5 ml of molten BBL top agar. The plate was incubated, usually at 37°C, until confluent lysis was achieved (after 4-6 h), when 3 ml of L-broth were added to the plate which was left at 4°C overnight. The broth was decanted, mixed with a few drops of chloroform and the debris pelleted by spinning in a bench centrifuge. The cleared supernatant was titrated and stored at 4°C.

2.2.c. Phage titration and spot tests

After serial dilution in phage buffer, 0.1 ml of phage suspension was mixed with 0.2 ml of plating cells, adsorbed for 10 min then plated out in 2.5 ml of molten BBL top agar on to a BBL plate. After overnight incubation at the appropriate temperature, usually 37°C,

the plaques were counted and the number of plaque forming units (pfu) per 1 ml of phage lysate calculated.

For spot tests on an appropriate indicator strain 0.01 ml aliquots of serial dilutions of phage were spotted on to a bacterial lawn from a 0.1 ml pipette. Spots were allowed to dry before the plate was incubated overnight at the appropriate temperature.

2.2.d. Phage crosses

2×10^8 plating bacteria and 2×10^9 phage (a total multiplicity of infection (moi) of 10 consisting of 5 of each phage) were mixed in less than 1 ml volume and adsorbed for 15 min. Cells were pelleted and resuspended in 1 ml of prewarmed L-broth before diluting 100-fold into warm broth. The culture was shaken at 37°C for 1½ h after which a few drops of chloroform were added. Phage were assayed on appropriate indicator bacteria to check for both parental phages and recombinants. Titres were usually 10^8 pfu/ml.

2.2.e. Construction of λ lysogens

Fresh plating cells were infected with phage at a moi of 1-2 and allowed to adsorb. The cells were diluted in L-broth and grown for 30-40 min at 37°C, or 32°C if the phage has a temperature-sensitive repressor. The cells were pelleted and resuspended in 10 m$\underline{M}$ $MgSO_4$ before plating out on L-agar plates in the presence of 10^9 pfu λ clear ($\underline{c}I^-$). Colonies which grew after overnight incubation were spot-tested for sensitivity to

λ vir. Titres were 10^8 pfu/ml.

2.2.f. T4 and λ phage liquid lysates

A fresh overnight of cells, usually C600 for λ and CR63 for T4, was diluted 50-fold into 200 ml L-broth (supplemented with 10 mM $MgSO_4$ for λ) in a 2 litre flask. The cells were shaken, usually at 37°C until they reached an OD_{650} of 0.45-0.6 (2-3 x 10^8 cells/ml). Phage were added to a moi of 0.1-1 for λ or 0.05 for T4 and incubation was continued at 37°C. The OD usually reached a maximum of 1.8 and then dropped as the cells lysed (after 2 h for T4 or up to 4 h for λ). When the OD reached a minimum, chloroform (1 ml/500 ml culture) was added and the flask shaken at 37°C for a further 15 min. The lysate was clarified by centrifugation, assayed at 10^{-7} and 10^{-8} dilutions, and checked on appropriate indicator strains. Titres of 5 x 10^{10} - 1 x 10^{11} were obtained for λ.

For λCM21 DNA preparations it was necessary to induce a lysogen, C600 (λCM21) (see Section u) and to concentrate the cells into a smaller volume before lysing with chloroform.

2.2.g. Preparation of CsCl purified phage and phage DNA

Phage in the cleared lysate (see above) were precipitated with polyethylene glycol (PEG) 6000. First 4 g/100 ml of NaCl were added followed by DNAse I and RNAse I, both to 1 µg/ml. After standing at room temperature for at least 1 h, 10 g of solid PEG 6000/100 ml

were added and dissolved by gently swirling. The flask was stood at 4°C for a minimum of 2 h, usually overnight. The PEG precipitate was recovered by centrifugation at 10,000 rpm for 10 min, and resuspended in 5 ml of phage buffer per 100 ml of starting volume by swirling gently at 4°C for 2-3 h. Debris was removed by centrifugation at 5,000 rpm for 10 min before concentration of the phage on a CsCl step gradient.

A step gradient was made in an MSE 14 ml polycarbonate tube; 1.5 ml of 31% w/w CsCl solution (3.1 g CsCl/6.9 ml phage buffer) was pipetted into the tube and underlayed with 1.5 ml of 45% w/w solution (4.5 g CsCl/5.5 ml phage buffer) using a pasteur pipette. Finally these two steps were underlayed with 1.5 ml of 56% w/w solution (5.6 g CsCl/4.4 ml phage buffer). The phage solution was overlayed on to the gradient and phage buffer was added to fill up the tube. Gradients were spun in a 6 x 14 Ti swing-out rotor in an MSE Superspeed 65 ultracentrifuge at 33,000 rpm for 2 h and stopped without braking. The phage band was collected in a 1 ml syringe by piercing the tube just below the band. Phage were stored in CsCl for long periods at 4°C.

Concentration on a step gradient provided sufficiently high titre lysates for infection experiments. Such lysates were dialysed for 2 h against phage buffer to remove CsCl and stored at 4°C, titres were about 10^{12} pfu/ml.

The phage from a step gradient also provided an adequate source of DNA for restriction enzyme analysis, however, where DNA was to be used as a cloning vector a further equilibrium gradient purification was carried out. The phage band was mixed with 41.5% w/w pre-clarified CsCl solution (20.75 g CsCl/29.25 ml phage buffer) in an MSE 5 ml polycarbonate tube and spun in a 6 x 5 Ti swing-out rotor in an MSE Superspeed 65 ultracentrifuge at 33,000 rpm for 24-36 h at 4°C. The phage band was collected in a syringe as before.

Before DNA was extracted the concentrated phages were dialysed against 10 mM Tris.HCl, pH 8.0, 1 mM EDTA for 2 h to remove CsCl. The phage protein was extracted 3 times with an equal volume of phenol pre-equilibrated with 0.5 M Tris.HCl pH 8.0, mixing gently by inversion in Eppendorf tubes or siliconised Corex tubes for large volumes. The phases were separated by spinning in a bench centrifuge and the lower phenol layer was removed using a drawn out pasteur pipette and discarded. The aqueous phase was extracted once with ether, or until it became clear. Finally, the DNA was dialysed against 10 mM Tris.HCl pH 8.0, 1 mM EDTA for 24 h with several buffer changes. The concentration of the DNA was determined by measuring the OD at 260 nm (an OD_{260} of 1 is equivalent to 50 µg/ml).

2.2.h. Preparation of M13 replicative form (RF) DNA

Early log phase cells, usually NM522 were prepared by diluting a fresh overnight 100-fold and growing to OD_{600} of 0.05-0.2. A single plaque was transferred to 2 ml of cells and the culture was shaken at 37°C for 12-18 h. 1.5 ml was transferred to an Eppendorf tube, spun for 5 min to pellet cells and the supernatant titrated. Titres were 10^{11} pfu/ml. 50 ml of early log phase cells were infected with phage at a final concentration of 10^9 pfu/ml and shaken at 37°C for 16-18 h. Cells were pelleted and the supernatant titrated to check that there were enough phage to infect a large culture.

An overnight culture was diluted 100-fold in 250 ml of L-broth in a 2 litre flask and grown at 37°C to OD_{600} of 1 (2-3 h). Phage were added to 10^{10} pfu/ml of cells and grown at 37°C for a further 2 h. Cells were sedimented by centrifugation at 3,500 rpm for 15 min at 4°C and the pellet resuspended in 50 ml of 50 mM Tris.HCl, pH 8.5, 1 mM EDTA before repelleting at 3,500 rpm for 15 min. The pellet was resuspended in 18 ml of 15% sucrose, 50 mM Tris.HCl pH 8.5, 50 mM EDTA. The suspension was kept on ice and pipetted gently to disperse before adding 4.5 ml of freshly prepared 5 mg/ml lysozyme in the above buffer. The cells were mixed gently while standing on ice for 10 min and were then lysed by addition of 13.5 ml of 0.1% Triton X-100, 50 mM Tris-HCl pH 8.5, 50 mM EDTA and gentle vortexing before

incubation at 37°C for 10-15 min. At this point the solution was very viscous.

Centrifugation at 12,500 rpm for 45 min at 4°C gave a greyish pellet and a clear supernatant. CsCl (0.95 g/ml) and ethidium bromide (0.59 mg/ml) were added to the supernatant which was loaded into MSE 10 ml polycarbonate tubes and overlayed with paraffin oil before caps were fitted. Gradients were generated by spinning the tube in a 10 x 10 Ti fixed angle rotor in an MSE Superspeed 65 ultracentrifuge at 45,000 rpm for 20 h at 20°C. Two bands were visible in UV light, the upper band consisting of nicked plasmid, linear plasmid and E.coli DNA while supercoiled closed circular plasmid M13 DNA formed the lower band. The lower band was collected in a syringe by piercing the side of the tube, and transferred to another 10 ml tube which was filled with a solution consisting of 1.059 g CsCl, 0.59 mg ethidium bromide per ml of distilled water. The tube was spun as described above, the single band collected in a volume of about 0.5 ml, and passed down a small 'Dowex' resin column made in a pasteur pipette plugged with siliconised glass wool. The DNA was eluted with 3 volumes of TE buffer pH 7.5. Remaining traces of ethidium bromide were removed by 5 or 6 extractions with an equal volume of a 1:1 ratio n-butanol/isopropanol solution (this solvent solution had been saturated with NaCl saturated TE buffer).

The DNA was dialysed overnight at 4°C against TE

buffer pH 7.5 to remove CsCl and was precipitated with 0.4 volume of 5 M ammonium acetate, 2 volumes of isopropanol and recovered by centrifugation. The final concentration of DNA calculated from the OD_{260} was 500-1000 µg per litre of culture.

2.2.i. Restriction enzyme digestion and ligation of DNA

DNA was digested with restriction endonucleases under conditions recommended by the suppliers, usually in a total volume of 10 µl. Reactions were stopped after 1-3 h of digestion by heating at 70°C for 10 min, or if the enzyme was not heat labile, particularly in the case of SmaI, by isopropanol precipitation. This was done by adding 0.4 volume of 5 M ammonium acetate and 2 volumes of isopropanol and standing at room temperature for 15 min before spinning at 10,000 rpm for 10 min to pellet the DNA. The pellet was washed with 1 ml of cold ethanol, repelleted at 10,000 rpm and dried in a vacuum desiccator before resuspending in an appropriate volume of TE buffer.

DNA was ligated using T4 DNA ligase usually in a 10 µl volume containing 50 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 1-100 units T4 DNA ligase (dependent on whether 'blunt' or 'sticky' ends were to be ligated). The ligation was incubated at 10°C for several hours, usually overnight. Recombinant phages were recovered by transfection or packaging.

2.2.j. Transfection and packaging

λ recombinants were recovered as plaques by transfection of cells made competent in the uptake of DNA by a modification of the method of Lederberg and Cohen (1974). A fresh overnight was diluted 50 fold into L-broth and grown at 37°C to an OD_{650} of 0.5. The cells were pelleted in a bench centrifuge at 4°C, washed in a half volume of ice cold 0.1 M $MgCl_2$, immediately repelleted and resuspended in a twentieth volume of ice cold 0.1 M $CaCl_2$. The cells were kept on ice for at least 30 min (up to several hours). DNA diluted in 0.1 ml of a 1 x SSC/0.1 M $CaCl_2$ (ratio 3:4) solution was added to 0.2 ml of competent cells and kept on ice for 30 min. After heating for 2 min at 42°C cells were returned to ice for a further 30 min before samples were plated out in 2.5 ml of molten BBL top agar on BBL plates and incubated overnight, usually at 37°C. Transfection efficiencies were about 10^6 pfu/µg DNA.

M13 recombinants were recovered by transfection of cells made competent by a modification of the method of Hanahan (1983). A fresh overnight of NM522 was diluted 500 fold and grown to an OD_{650} of 0.6. The cells were stood on ice for 15 min before pelleting and resuspending in a third volume of TFB buffer. The cells were kept on ice for 15 min, repelleted and resuspended in a twelfth volume of TBE buffer. While cells were kept on ice 7 µl DMSO per 200 µl cells were added. After 5 min 7 µl 2.5 M DTT per 200 µl cells

were added and after a further 10 min, 7 µl DMSO per 200 µl cells. After 5 min 200 µl of cells were added to the DNA in chilled Eppendorf tubes which were kept on ice for 45 min before heating at 42°C for 1½ min and plating out aliquots in 2.5 ml BBL top agar containing 0.2 ml NM522 plating cells and indicator for β-galactosidase (see below) on minimal plates. Transfection efficiencies were 10^7 pfu/µg.

Packaging mixes for λ were supplied by A. J. B. Campbell. The freeze thaw lysate (FTL) and sonicated extract (SE) were thawed on ice and added to incubation buffers in the following order: 7 µl buffer A, 1-2 µl λ DNA (between 0.2-0.5 µg), 1 µl buffer M1, 6 µl SE and 10 µl FTL. The reaction was incubated for 60 min at 25°C, and diluted with 500 µl of phage buffer. Aliquots were plated out on BBL plates with appropriate indicator bacteria.

2.2.k. Selection of recombinants

Recombinants of λ immunity vectors e.g. NM1149 were recovered on a hfl$^-$ E.coli strain, NM514 which only allows cI$^-$ phage to form plaques and hence selects for λ molecules containing an insert in the cI gene. Those non-recombinant vector molecules still able to express cI repressor are repressed and unable to propagate lytically.

λ recombinants with amber mutations were recovered on appropriate suppressing strains and identified

by their inability to grow on a sup° host.

M13 recombinants were recognised as white plaques recovered by plating in BBL top agar containing 0.2 mg/ml of indicator (5-bromo-4-chloro-3 indolyl-β-D-galactoside, Xg) and 0.25 mg/ml of inducer (isopropyl-β-D-thiogalactosidase, IPTG). Non-recombinant M13 vector molecules formed blue plaques.

2.2.1. Preparation of filters for screening by hybridisation

Phage recombinants were detected by probing denatured phages baked on to nitrocellulose filters with appropriate radiolabelled DNA probes.

Plaques were blotted from plates either directly or after transferring by toothpick to a lawn of bacteria marked out as a grid. Appropriate phage were included as positive and negative controls. Use of dry BBL plates helped to prevent tearing of the top agar layer during blotting. Circular nitrocellulose filters were cut to size and the orientation marked by notches or ink. A filter was placed on a cooled plate for 1 min and then placed, plaque side up on blotting paper soaked in denaturation buffer for 2 min. The filter was transferred to blotting paper soaked in neutralisation buffer for a few minutes before a final wash in a tray of 2 x SSC. The filter was blotted dry and baked at 80°C under vacuum for 2 h and stored at 4°C.

Where the probe to be used was single-stranded M13 DNA, it was necessary to hybridise to single-stranded

M13 DNA derived from phage lysates since cells within a plaque contain double-stranded RF DNA which will hybridise to the vector. A single plaque was picked into 1.5 ml of a 100-fold dilution of an NM522 overnight and grown for 4½-5½ h at 37°C before pelleting cell debris. 3 µl of the cleared supernatant was spotted on to a nitrocellulose filter marked out as a grid. The filter was treated and baked as described above.

2.2.m. Radiolabelling of double-stranded probes by nick-translation

Deoxycytidine 5'-[α-^{32}P]triphosphate (10 µCi) was dried down under vacuum and resuspended in 20 µl of 1 x dNTP buffer, 1 µl DNAse I (at 2 x 10^{-5} mg/ml), 1 µl DNA polymerase I (at 1 unit/µl) and 0.25-1.0 µg DNA (in $\sim$ 2 µl). After incubation for 1-3 h at 15°C the reaction was stopped by the addition of 200 µl of TE pH 8.0 and loaded on to a Sephadex G-50 column. The labelled DNA was eluted with TE buffer and collected as the first peak in a volume of 0.5-1.0 ml. The amount of label incorporated was determined with a liquid scintillation counter (5 µl samples on Whatman GF/C disks were dried and counted in a non-aqueous scintillant).

Filters were prehybridised for 30 min at 37°C in 30 ml of hybridisation buffer. The probe ($\sim 10^6$ counts per filter) was added to 250 µg of sonicated calf thymus DNA, denatured at 95°C for 10 min and immediately cooled on ice. The probe was added to the prehybridised

filter with 10-15 ml of fresh hybridisation buffer and the hybridisation was carried out at 37°C overnight with gentle shaking. The filter was washed twice in 2 x SSC, 0.1% SDS for 30 min at 37°C, and twice in 1 x SSC, 0.1% SDS for 30 min at room temperature. Hybridisation of the probe to the filter was detected by autoradiography at -70°C using pre-flashed X-ray film.

2.2.n. Radiolabelling of single-stranded M13 DNA

1 µl of M13 hybridisation probe primer was boiled for 1 min, cooled on ice then added to 2 µl (2 ng) M13 single-stranded template DNA (see section q(ii)), 1 µl 10 x HincII buffer, 1 µl 0.1 M DTT, 5 µl H_2O. This annealing mixture was incubated at 65°C for 15 min and cooled to room temperature. Deoxycytidine 5'-[α-^{32}P]triphosphate (10 µCi) was dried under vacuum and dissolved in the annealing mixture. Cold dNTPs were added as 1 µl of a solution containing dGTP, dATP, dTTP all at 500 µM together with 0.5 unit of Klenow polymerase. The reaction was incubated at 15°C for 90 min and the enzyme inactivated by addition of 1 µl of 250 mM EDTA pH 8.3. This probe was kept on ice before use and not denatured.

Filters were prehybridised at 65°C for several hours in a solution containing 5 x Denhardt solution, 5 x SSC, 50 µg/ml denatured sonicated calf thymus DNA, 0.1% SDS. The probe was added to the prehybridised filter in 10-15 ml of a solution containing 1 x Denhardt

solution, 5 x SSC, 50 µg/ml denatured sonicated calf thymus DNA, 0.1% SDS. Hybridisation was carried out overnight at 65°C with gently shaking. Filters were washed 4 times for 1 h in 0.5 x SSC, 0.1% SDS at 65°C with shaking.

2.2.o. Agarose gel electrophoresis of DNA

Fragments from restriction endonuclease digests of DNA were analysed by separation on 0.7-0.8% agarose (Miles) gels made up in TBE buffer.

Samples of DNA, usually 0.2-0.5 µg in a 0.5 cm slot, were loaded in 1 x TBE loading dye. λ DNA was always heated at 65°C for a few minutes to melt the cohesive ends immediately before loading. Unsubmerged gels were run overnight at 15-20 mA and stained for 40 min in a 1 µg/ml solution of ethidium bromide. After destaining for 30 min in distilled water the gel was photographed over a short wave, UV light transilluminator.

Mini-gels poured in small gel tanks or on microscope slides were often sufficient for checking restriction digests. These could be run in about 1 h at 100 V in TBE buffer, stained for a few minutes in ethidium bromide and the DNA visualised under UV light.

2.2.p. Isolation of DNA fragments from agarose and polyacrylamide gels

DNA fragments larger than 1 kb were isolated from agarose (Seakem) gels. The gel was stained with ethidium bromide and the DNA visualised on a low power

UV transilluminator to minimise damage to the DNA. The appropriate band was excised from the gel, chopped into small pieces and loaded into dialysis tubing with 1-2 ml of 0.5 x TBE buffer. The tubing was sealed after removing trapped air, and immersed in a shallow layer of 0.5 x TBE in an electrophoresis tank. An electric current of 100 V was passed through the bag for 2-3 h, and eventually the current was reversed for 1 min to release DNA stuck to the wall of the dialysis tubing. The buffer was recovered from the dialysis tubing the inside of which was rinsed out with 1 ml 1 x TBE buffer. The two solutions were pooled and the DNA was recovered on a DEAE cellulose (DE52) column made in a Gilson p1000 plastic tip plugged with siliconised glass wool. 0.6 ml of resin (sufficient to bind 20 µg of DNA) was packed into the column and washed with 3 ml TE buffer pH 7.6 containing 0.6 M NaCl, 3 ml TE pH 7.6 alone, 3 ml TE pH 7.6 containing 0.1 M NaCl. The DNA in gel buffer was then loaded on to the column and the flow-through reapplied. The column was washed twice with 1.5 ml of TE containing 0.3 M NaCl, and the DNA eluted with three 0.5 ml washes of TE containing 0.6 M NaCl. The eluate was precipitated with 2.5 volumes of ethanol at -20°C overnight, adding tRNA as carrier where necessary*. The DNA was pelleted by spinning for 10-30 min at 10,000 rpm, washed once with 95% ethanol and repelleted. The pellet was dried under vacuum and resuspended in a small volume of TE buffer.

* 0.1 to 1 µg of DNA in 1.5 ml was recovered at good efficiency by the addition of ~ 20 µg/ml carrier tRNA.

DNA fragments of less than 1 kb were isolated from polyacrylamide gels essentially by the method of Maxam and Gilbert (1977). A polyacrylamide gel of an appropriate concentration for maximum separation (Maniatis et al., 1982) was made up in TBE buffer from a 30% stock solution of acrylamide (see solutions) and cast between vertical gel plates separated by spacers. A slot former was inserted and the acrylamide allowed to polymerise for 1 h. The gel was then attached to an electrophoresis tank filled with 1 x TBE buffer and up to 1 µg of DNA was loaded into 0.5 x 0.2 cm wells in 1 x TBE loading dye. The gel was run for several hours at between 1 and 8 V/cm ensuring that it did not overheat and thereby denature the DNA. The gel was stained for 40 min in a solution of 0.5 µg/ml of ethidium bromide in 1 x TBE buffer. The DNA was visualised over a UV transilluminator and the appropriate band excised. Polyacrylamide quenches fluorescence and more than 10 ng of DNA in a band is necessary for detection. The gel slice was chopped into small pieces and transferred to a siliconised test tube containing 1 volume of elution buffer (0.5 M ammonium acetate, 1 mM EDTA). The tube was sealed with parafilm and incubated at 37°C overnight on a rotating wheel. The fragments of acrylamide were removed by filtering through siliconised glass wool held within an Eppendorf tube from which the tip had been removed. This Eppendorf was supported in a small siliconised glass tube which was spun at 3,000 rpm

in a bench centrifuge for 5 min so that the cleared solution collected in the glass tube after passing through the glass wool. The glass wool was rinsed once with a half volume of elution buffer in the same way. These solutions were pooled and extracted 3 or 4 times with butanol saturated in TE buffer to remove ethidium bromide. Finally the solution was extracted once with phenol, precipitated twice with 0.1 volume 3M sodium acetate, 2.5 volumes of ethanol, washed with 70% ethanol and dried under vacuum. The DNA was resuspended in a small volume of TE buffer.

2.2.q. Dideoxy chain termination sequencing of DNA fragments cloned in M13.

(i) Generation and cloning of random DNA Fragments

About 10 µg of purified DNA fragment was circularised by ligating the 'sticky' ends in the presence of 20 units of T4 DNA ligase in a total volume of 30 µl. The ligated DNA in an Eppendorf tube was sonicated at maximum power for 4 bursts of 40 sec in a cup horn sonicator (Heat Systems Ultrasonics, Inc. Model W-375). The tube was spun briefly after each burst to 'concentrate' the solution in the bottom of the tube. The ends of the DNA fragments were repaired to provide blunt ends for ligation to SmaI cut M13 vector DNA. To 28 µl of the

sonicated DNA suspension were added 2 µl of sequence chase mix, 3 µl of TM buffer and 20 units of Klenow polymerase. THE DNA was incubated at 15°C overnight and run on an agarose mini-gel next to a Sau3a digest of pAT153 which provided size markers. The smear of DNA in the size region 300-600 bases was excised from the gel, recovered by electroelution in dialysis tubing (section 2.2.p) and dissolved in about 50 µl of TE buffer. 1-5 µl aliquots of the end repaired fragments were ligated to 20 ng SmaI cut M13 vector DNA and recombinants were recovered by transfection of NM522 cells.

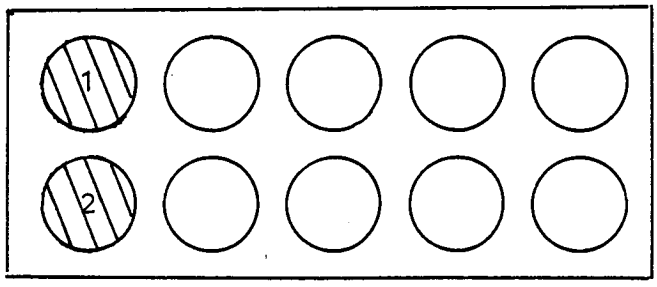
(ii) Growth of recombinant M13 phage and extraction of single-stranded template DNA

A single white plaque from the transfection plate was picked into 50 µl of phage buffer and then serially diluted to 10^{-6}. 0.1 ml of this dilution was plated out with 0.2 ml of NM522 plating cells in the presence of Xg and IPTG. A white plaque from this plate was picked into 1.5 ml of a 100-fold dilution of an NM522 overnight in a small sterile bottle. The culture was grown with vigorous shaking at 37°C for 4.5-5.5 h, and then transferred to an Eppendorf tube and clarified by spinning at 10,000 rpm for 5 min. The supernatant was

transferred to a clean tube and 150 µl of 20% PEG 6000, 2.5 M NaCl solution was added. The tube was kept at 4°C for 15 min before pelleting the phage by spinning at 10,000 rpm for 10 min. The supernatant was discarded, the tube spun briefly and all residual PEG solution removed with a drawn out pasteur pipette. The pellet was resuspended in 100 µl of TE buffer and extracted with 50 µl of TE equilibrated phenol. Finally 0.1 volume of 3 M sodium acetate and 2.5 volumes of ethanol were added and the DNA was precipitated overnight at -20°C. The DNA was pelleted by spinning at 10,000 rpm for 10 min, washed with 1 ml of cold ethanol and repelleted. It was finally dissolved in 30 µl of TE buffer and stored at -20°C.

(iii) Dideoxy chain termination sequencing

The template was annealed to M13 17-mer sequencing primer. The primer mix consisting of 2 µl 17-mer primer (0.4 pmol), 1 µl TM buffer, 2 µl H_2O was added to 5 µl of template DNA and annealed at 60°C for 1 hour. The mix was allowed to cool slowly to room temperature before 2 µl was dispensed into each of 4 capless tubes in a centrifuge rack:



clone	T	C	G	A	
template/primer	2	2	2	2	µl
T mix	2				
C mix		2			
G mix			2		
A mix				2	
Klenow mix	2	2	2	2	

The appropriate termination mix was added to each tube (see Solutions) and finally 2 µl of Klenow polymerase mix was added to each tube. The drops were deposited on the sides of the tubes so that all reactions could be started simultaneously by a brief spin in an Eppendorf centrifuge (model 5413). After 25 min at room temperature 2 µl of sequencing chase mix was added to each tube, the racks spun again and incubated for a further 15 min. At this point reactions were either stored at -20°C or 2 µl of formamide dyes were added to each tube and the reactions boiled for 3½ min before loading on to a separating gel.

(iv) Buffer gradient polyacrylamide gel electrophoresis

The use of a crude buffer gradient gel (Biggin et al. 1983) allowed at least 250 bases to be read from each clone. For each gel the following were prepared:

(1) 7 ml 2.5 x TBE gel mix + 14 µl 25% AMPS + 14 µl TEMED

(2) 35 ml 0.5 x TBE gel mix + 70 µl 25% AMPS + 70 µl TEMED

Using a 10 ml pipette, 4 ml of 0.5 x TBE gel mix (2) then 6 ml of 2.5 x TBE gel mix (1) were taken up, mixed by pulling through an air bubble, and poured down between the vertical gel plates separated by 0.35 mm "plasticard" spacers. The flow into the plates was continued with the remainder of the 0.5 x TBE gel mix (2) from a 30 ml syringe until plates were full. A "plasticard" slot former with 3 mm wide wells, 1.5 mm apart was set between the plates which were clamped firmly and the gel was allowed to polymerise for 1 h. Sequencing reactions were boiled and samples loaded in order using a drawn out capillary tube. The gel was run at 25-30 watts for about 2 h in 1 x TBE buffer. For clones with inserts in the SmaI site of M13 the bromophenol blue marker was allowed to run off, and after 15 min the gel was removed. The gel

was left attached to the unnotched plate for support and fixed for 10 min in 10% acetic acid, 10% methanol solution. It was then drained, transferred to damp blotting paper and covered with Saranwrap plastic film. After drying on a vacuum gel dryer at 80°C the Saranwrap was removed and the gel placed in direct contact with X-ray film and left to autoradiograph overnight at room temperature.

Gel readings were compiled and analysed on a VAX/VMS computer using the programmes of Staden (1982) and the University of Wisconsin Genetics Computer Group.

2.2.r. Labelling of polypeptides following infection of UV irradiated cells

An overnight of M159 cells grown in M9-maltose medium was diluted with fresh medium to an OD_{650} of 0.1 and grown at 37°C to an OD_{650} of 0.5 (4-5 h). 15 ml of cells were pelleted in a bench top centrifuge, the supernatant removed carefully and the cells resuspended in 4.5 ml of ice-cold M9-maltose medium supplemented with 20 m$\underline{M}$ $MgSO_4$. 4 ml of the suspension was UV irradiated in an open petri dish for 7½ min at 4,500 erg/mm^2 with constant gentle swirling. The irradiated cells were transferred to a foil-wrapped flask to exclude visible light and stood on ice for 10 min.

CsCl purified phage which had been dialysed

against phage buffer were diluted to 10^{11} pfu/ml and 5 µl (5×10^8 pfu) placed in snap cap Eppendorf tubes on ice. 50 µl (5×10^7) of irradiated cells were added and the phage adsorbed on ice for 10 min. The tube was warmed at 37°C for 2 min before the addition of 200 µl of pre-warmed M9-maltose medium supplemented with 0.04 µg/ml of cold L-methionine. 20 µCi of L-[^{35}S]methionine were added 3 min later to label proteins synthesised early after infection by λ, or 20 min later to label proteins synthesised late in infection. For labelling T4 proteins, label was added after 4 min. The tubes were incubated at 37°C for 10 min before adding 50 µl of 1 mg/ml cold L-methionine. After a further 10 min at 37°C the tubes were put on ice for a few minutes before pelleting the cells at 10,000 rpm for 10 min at 4°C. The supernatant was removed with a drawn out pasteur pipette and the pellet resuspended in 1 ml of ice-cold acetone, repelleted at 10,000 rpm and finally resuspended in 100 µl of sample buffer. Samples were boiled for 2 min and the amount of label incorporated was checked by counting a 5 µl sample in a scintillation counter. Samples could be stored frozen at -70°C for a few weeks.

2.2.s. <u>SDS polyacrylamide gel electrophoresis of polypeptides</u>

36 x 25 cm gels were poured between glass plates separated by 1 mm spacers and sealed together with 1 mm bore silicone rubber tubing and bulldog clips. The appropriate quantity of acrylamide gel solution was made

up as described without TEMED which was only added just before pouring. The ratio of bis acrylamide to acrylamide in the stock solution was chosen to give the optimum separation of polypeptides in the 30-40K region (Hancock et al., 1976). The separating gel was poured directly between the plates to within 4 cm of the top and overlayed with water saturated butanol. After 1 h the butanol was rinsed off first with ethanol and then water, and a slot former was inserted between the plates. The stacking gel was poured and allowed to polymerise for 1 h. The slot former, tubing and bottom spacer were removed and the gel clamped on to the electrophoresis tank which was filled with 1 x Laemmli electrophoresis buffer. The wells were thoroughly washed out and the protein samples were loaded with a microsyringe after boiling for 2 min. Usually 0.2-0.5 mg of protein, or 2-5 x 10^5 counts were loaded per well. The gel was run at 20-24 mA overnight.

2.2.t. Staining and autoradiography of protein gels

Where mg quantities of protein were present gels were stained with Coomassie blue (0.05% w/v in 10% acetic acid, 25% methanol) for 2 h and destained over a few hours with several changes of 10% acetic acid, 25% methanol.

Silver staining was used for smaller quantities of protein. The gel was gently shaken throughout at room temperature; the gel was first washed for 30 min in 45% methanol, 10% acetic acid, prefixed for 30 min in 5% methanol, 7% acetic acid and fixed for 30 min in

10% glutaraldehyde. It was then rinsed thoroughly in distilled water with repeated changes for over 1 h or overnight, then washed for 30 min in 5 µg/ml DTT, and 30 min in 0.1% $AgNO_3$. The gel was rinsed very quickly in distilled water and twice in developer (3% Na_2CO_3) and finally developed in 3% Na_2CO_3 containing 100 µl of 37% formaldehyde (HCOOH) per 100 ml of solution. After developing to the required intensity 5 ml of 2.3 M citric acid were added, the gel washed thoroughly in distilled water and stored sealed in a plastic bag with 0.03% Na_2CO_3.

Where radiolabelled proteins were used the gel was fixed for 2 h in 10% acetic acid, 25% methanol then dried on to blotting paper in a heated Bio-Rad Slab gel dryer (model 224) and autoradiographed by exposure to X-ray film at room temperature for 1-4 days.

2.2.u. Phage infection and prophage induction to obtain pnk expression

(i) Infection

ED8689 (sup°) cells were grown at 37°C to an OD_{650} of 0.4 ($\sim 2 \times 10^8$ cells/ml) and phage were added at a moi of 5 (or as high as possible). Where T4 or λ phage capable of cell lysis e.g. λCM8 were used, infection was allowed to proceed at 37°C for 10-15 min. Where phage had amber mutations preventing cell lysis e.g. λCM21, infection proceeded for 2-3 h. Cells were pelleted and supernatants discarded. The pellets were stored

at -20°C or were resuspended in 50 mM Tris.HCl pH 7.4, 1 mM 2-mercaptoethanol for cell disruption by sonication.

(ii) Induction

An overnight of a lysogen grown at 32°C was diluted 50 fold and grown at 32°C to an OD_{650} of 0.5. The cells were pelleted and resuspended in 1/5 volume of L-broth. The concentrated cells were heated at 42°C for 10 min in a shaking water bath, diluted to the original volume with L-broth at 37°C and grown at 37°C for 2-3 h before pelleting as described above.

2.2.v. Purification of polynucleotide kinase protein

The procedure was basically that described by Richardson (1965). All steps were carried out at 4°C unless otherwise stated, all centrifugations were at 10,000 rpm for 30 min at 4°C.

(i) Cell disruption

2-3 g of cell paste derived from a cell infection or prophage induction were mixed with 10 ml of 50 mM Tris.HCl pH 7.4, 1 mM 2-mercaptoethanol. The suspension was sonicated 8 times for 30 sec with constant cooling on ice and debris was removed by centrifugation.

(ii) Streptomycin precipitation

The supernatant was made 0.8% in streptomycin sulphate (2 ml of 5% streptomycin solution in 10 ml)

by dropwise addition over 30 min. After stirring for a further 15 min, or overnight, the suspension was sedimented by centrifugation and the supernatant discarded. The pellet was resuspended in 12 ml of 0.1 M potassium phosphate buffer (KPO_4) pH 7.5, 0.01 M 2-mercaptoethanol.

(iii) Autolysis

The suspension was incubated at 37°C for 1½ h with stirring then 0.0336 ml of 1 M $MgCl_2$ was added and incubation at 37°C continued for 1 h. The autolysate was chilled on ice for 1 h and sedimented by centrifugation.

(iv) Ammonium sulphate fractionation

The supernatant was made 10% in ammonium sulphate by gradual addition of well ground solid (1.2 g/12 ml) over 30 min with constant stirring. After a further 1 h of stirring the suspension was sedimented by centrifugation and the supernatant transferred to a clean tube and a further 20% of ammonium sulphate added (2.4 g/12 ml) over 30 min. After a further 1 h the suspension was sedimented, the supernatant discarded, and the pellet dissolved in 10 ml of 0.1 M KPO_4 pH 7.5, 0.01 M 2-mercaptoethanol. This suspension was dialysed against 0.01 M KPO_4 pH 7.5, 0.01 M 2-mercaptoethanol for 5 h.

(v) DEAE-Sephadex fractionation

A 1.5 x 10 cm glass column was packed with DEAE-Sephadex (A50) in equilibrating buffer (0.01 M KPO_4 pH 7.5, 0.01 M 2-mercaptoethanol). The column was washed with ~300 ml of equilibrating buffer and the dialysed protein suspension applied. The column was washed with another 100 ml of equilibrating buffer and the protein was eluted with equilibrating buffer made 0.1 M in KCl. 5 ml fractions were collected and assayed for kinase activity or, where the defective protein was purified, aliquots of fractions were analysed on SDS-PAGE by silver staining. Appropriate fractions were pooled and dialysed against 0.05 M KPO_4 pH 7.5, 0.01 M 2-mercaptoethanol.

(vi) Phosphocellulose fractionation

A 1 x 10 cm glass column was packed with phosphocellulose (P11) in equilibrating buffer (0.05 M KPO_4, pH 7.5, 0.01 M 2-mercaptoethanol) and washed with 300 ml of equilibrating buffer. The dialysed DEAE-Sephadex fractions were loaded on to the column which was washed with a further 10 ml of equilibrating buffer. The protein was eluted with 20 ml of equilibrating buffer made 0.05 M in KCl, 20 ml of buffer 0.1 M in KCl, 20 ml of buffer 0.25 M in KCl and 20 ml of buffer 0.5 M in KCl. Ten 2 ml fractions were collected at each step of the elution and these were again

assayed for kinase activity or analysed on silver stained SDS-PAGE.

Fractions containing enzyme were pooled and dialysed overnight at 4°C against 50% glycerol (Analar grade), 25 mM Tris.HCl pH 7.5, 10 mM 2-mercaptoethanol. The enzyme could then be stored at -20°C and retained activity for some months.

2.2.w. Lowry-Folin Assays for Protein Concentration

Protein concentration was assayed by the method of Lowry et al. (1951). Infected cells or induced lysogens were disrupted by sonication and pelleted to remove debris. 1, 3 and 5 µl aliquots of the cleared supernatant were diluted in 0.4 ml distilled water and 2 ml of reagent C added to each sample. The solution was mixed immediately and stood for 10 min before the addition of 0.5 ml of reagent D. After mixing the solution was stood for 30 min before the OD_{650} was measured. The OD reading was related to protein concentration by reference to a standard curve derived from a series of bovine serum albumin samples in the range 0-200 µg/ml.

2.2.x. Assay for kinase activity

The assay was based on the conversion of a dinucleoside monophosphate, cytidylyl-(3'-5')-uridine (CpU), to the dinucleotide (pCpU) by the incorporation of a labelled phosphate from [γ-^{32}P]ATP (Murray, 1973).

This quick simple assay was ideal for analysis of large numbers of crude cell extracts or fractions from pnk purification. However, analysis of crude extracts prepared by disruption of cells by sonication showed that kinase activity was manifest as the appearance of radioactive pC due to the action of a nuclease upon the primary product, labelled pCpU.

Cell pellets derived from 25-50 ml of culture were resuspended in 0.4 ml of 50 mM Tris.HCl, pH 7.4, 1 mM 2-mercaptoethanol in Eppendorf tubes and sonicated in 6 bursts of 3 seconds using a Dawe Soniprobe sonicator (model 7540A) ensuring that the tubes always remained cool. Debris was removed by centrifugation. 2.5 µl of clear supernatant (or fraction from the purification procedure) was incubated with 25 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.1 mM ATP, 1 mg/ml CpU, 0.1-1 µCi [γ-^{32}P]ATP in a 10 µl volume at 37°C for 30 min. The products were analysed by ionophoresis on AE-cellulose paper at pH 3.5.

2.2.y. Preparation of substrates, and assays for 5'-kinase and 3'-phosphatase specificities of purified pnk

(i) 5'-kinase assay

5'-hydroxyl deoxyribooligonucleotides were obtained from a digest of salmon sperm DNA with pancreatic DNAse (DNAse I) and calf intestinal phosphatase (CIP). Salmon sperm DNA was dissolved to 1.5 mg/ml in TE buffer pH 7.5. 10 µg

of the DNA was incubated in 0.01 M Tris.HCl pH 7.5, 0.01 M $MgCl_2$, 0.1 mg/ml DNAse I, 0.1 mg/ml CIP in a 100 µl volume at 37°C for 3 h . The reaction was heated at 70°C for 10 min.

5'-hydroxyl ribooligonucleotides were obtained from digestion of yeast tRNA with RNAse T_1. 5 µg of tRNA were incubated with 0.01 M Tris.HCl pH 7.5 and 20 units of RNAse T_1 in a 50 µl volume at 37°C for 2 h.

Aliquots of 5'-hydroxyl substrates were incubated with pnk under the conditions described in Section x. These phosphorylated oligonucleotides were further digested to single nucleotides with venom phosphodiesterase to investigate the specificity of labelled phosphate incorporation. A 10 µl aliquot of kinase treated oligonucleotides was incubated with 0.01 M Tris pH 8.5, 0.1 mg/ml venom phosphodiesterase at 37°C for 40 min.

(ii) 3'-phosphatase assay

5'-[^{32}P]dCMP and 3'-[^{32}P]dCMP, the substrates for this assay were isolated from appropriate digests of nick-translated λ cI857 DNA (i.e. labelled with [α-^{32}P]dCTP). 5'-[^{32}P]dCMP was obtained from sequential digestion of the labelled DNA with pancreatic DNAse and venom phosphodiesterase. An aliquot of DNA was

incubated with 0.01 M Tris.HCl pH 7.5, 0.01 M $MgCl_2$, 0.1 mg/ml DNAse I in a 50 µl volume at 37°C for 1 h. An aliquot of this digest was incubated with 0.01 M Tris.HCl pH 8.5, 0.1 mg/ml venom phosphodiesterase in a 50 µl volume at 37°C for 40 min.

3'-[^{32}P]dCMP was obtained from a sequential digest of labelled DNA with micrococcal nuclease and spleen phosphodiesterase. An aliquot of DNA was incubated with 0.3 M glycine buffer pH 9.2, 0.1 M $CaCl_2$, 0.3 mg/ml micrococcal nuclease in a 30 µl volume at 37°C for 1 h. The pH of the reaction was reduced to 6 by addition of acetic acid then an aliquot was incubated with 0.1 M sodium succinate . HCl pH 6.5, 1 unit/ml spleen phosphodiesterase in a 100 µl volume at 37°C for 1 h.

The reaction products were separated by ionophoresis on AE81 paper at pH 3.5 and the labelled spots located by autoradiography with X-ray film at -70°C. The 3'-[^{32}P]CMP and 5'-[^{32}P]CMP spots were cut out and eluted from the paper and incubated with 0.06 M sodium succinate.HCl pH 6.5, 0.01 M $MgCl_2$, 0.01 M 2-mercaptoethanol and enzyme at 37°C for 1 hr.

The products of all assays were analysed by ionophoresis on AE81 paper at pH 3.5.

2.2.z. High voltage paper ionophoresis, and solvent elution

Separation of components of digests and assays was achieved by high-voltage ionophoresis on AE-cellulose (AE81) paper at pH 3.5 (Sanger et al., 1965; Murray, 1970). 5 µl aliquots of the samples were spotted on to AE81 paper, 9 cm from one end and 0.5 cm apart (leaving a 4 cm gap down the sides of the paper where a marker dye could be loaded). A mark was made 23 cm from the origin and when spots were dry the paper was soaked in 0.5 x pH 3.5 buffer (1 x buffer is 1% pyridine, 10% acetic acid). The paper was lowered into the electrophoresis tank and run in 1 x pH 3.5 buffer at 3 kV. The ATP remained close to the origin whereas the four mononucleotides were well resolved when the fast xylene cyanol marker dye (equal amounts of xylene cyanol FF, orange G, and acid fuchsin in distilled water) had migrated 23 cm. After drying the paper, radiolabelled spots were detected by autoradiography with X-ray film at -70°C. Non-radioactive mononucleotide markers were visualised under UV light.

Separated spots could be eluted from the paper with alkaline triethylamine carbonate. (CO_2 was passed into a mixture of 70 ml of water and 30 ml of redistilled triethylamine until saturated. The pH was adjusted to 10 with triethylamine.) The radiolabelled spot was cut out as a rectangle with one pointed end which was rested against a drawn out capillary tube while the other

end was sandwiched between glass slides standing in a trough of triethylamine. Capillary action draws solvent into the capillary from the paper. The solvent, containing the radiolabelled nucleotide species was spotted on to polythene and dried under vacuum without allowing the solvent to bubble. The spot was dissolved in water and redried 3 times before finally resuspending in 0.01 M Tris.HCl pH 8.5 and digesting with venom phosphodiesterase.

CHAPTER 3

RESULTS

3.1 CLONING AND EXPRESSION OF THE pseT GENE

The location of pseT in the T4 genome is inferred by the absence of T4 EcoRI fragments 21 (2.27 kb), 33 (1.1 kb) and 46 (0.1 kb) (Figure 3.1) from the kinase deficient deletion strain T4pseTΔ1 (Mileham et al., 1980). An analysis of heteroduplex molecules indicates that pseTΔ1 is 3.5 kb in length (Kutter et al., 1984) so it is unlikely that pseTΔ1 extends very far outside of EcoRI fragments 21, 33 and 46. An open reading frame (ORF) of about 1000 bases, sufficient to encode a 33 kd polypeptide, could be contained entirely within the 2.27 kb or the 1.1 kb EcoRI fragments. The initial approach therefore was to analyse polypeptides expressed by the larger EcoRI fragments, and to determine the DNA sequence with the intention of identifying the pseT gene.

3.1.a. Correlation of a T4 33 kd polypeptide with pnk activity in infected cells

UV irradiated cells (M159) were infected with T4 phage (see Section 2.2.r) and polypeptides labelled with [^{35}S]methionine were separated on 15% linear SDS polyacrylamide gels and detected by autoradiography (Figure 3.2). In infections with phage known to express an active pnk (wild-type T4D or T4amN122) a 33 kd polypeptide was present, whereas this polypeptide was missing in infections with the pseT$^-$ strain T4pseT2 (Figure 3.2). In

Figure 3.1

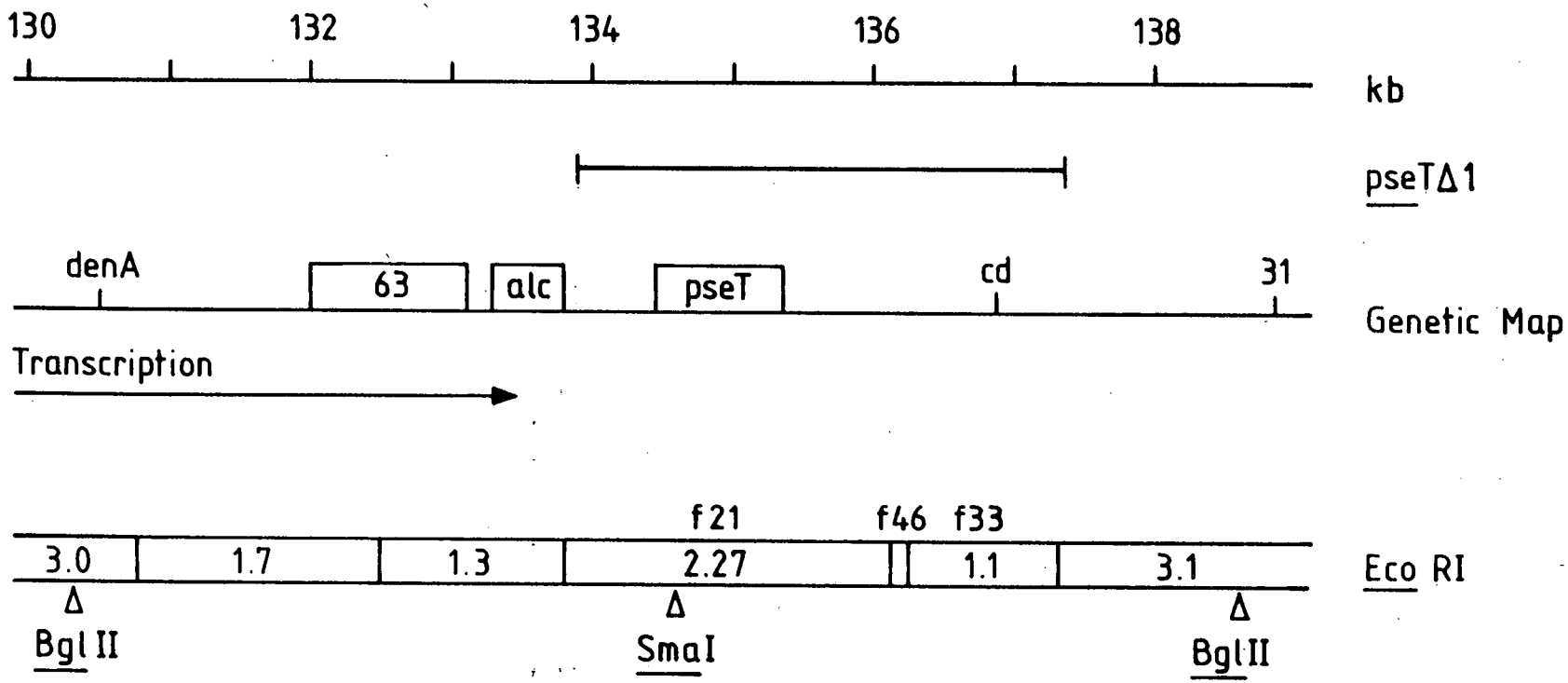

130
132
134
136
138
kb
pseTΔ1
denA
63
alc
pseT
cd
31
Genetic Map
Transcription
f21
f46
f33
3.0
1.7
1.3
2.27
1.1
3.1
Eco RI
Δ
Bgl II
Δ
SmaI
Δ
BglII

Figure 3.1

Organisation of the pseT region of the T4 genome. The figure shows: physical distance in kilobase pairs from the rIIA-rIIB join = 0 (Wood and Revel, 1976); the map position of T4 deletion pseTΔ1 (Mileham et al. 1980; Kutter et al., 1984); a genetic map of T4 showing gene order, rectangles represent genes whose lengths have been determined by DNA sequence; the direction of transcription; the numbers alloted to the relevant EcoRI fragments (Kutter et al., 1984); restriction sites for EcoRI, BglII and SmaI.

addition to the correlation between the presence of a 33 kd polypeptide and kinase activity, a smaller polypeptide of about 20 kd exhibited different mobilities when infections with the pseT$^+$ strains T4D or T4amN122 were compared with a T4pseT2 (pseT$^-$) infection. The T4 pseT2 and T4amN122 strains were therefore crossed and a sample of amber recombinants unable to form plaques on a suppressor-free strain (ED8689) were assayed for induction of kinase activity on infection of non-irradiated cells (ED8689) (Figure 3.3). Recombinants with the amN122 mutation in T4 gene 42 (dCMP-hydroxymethylase) are defective in DNA synthesis. Since DNA synthesis is involved in switching off early genes, amN122 mutants continue to synthesise pnk throughout infection which improves the yield. The infected cells were sonicated to release cell proteins and samples of the cleared crude cell lysates were assayed for kinase activity by their ability to convert a dinucleoside monophosphate CpU, to the dinucleotide pCpU, by the incorporation of labelled phosphate from [γ-^{32}P]ATP. Products of the reactions were separated by ionophoresis at pH 3.5 on AE81 cellulose paper. In Figure 3.3, [^{32}P]pCpU is only seen as the product of an assay using purified T4 encoded pnk. In assays of crude cell extracts, a nuclease degrades [^{32}P]pCpU to [^{32}P]pC. This nuclease activity is detected when purified pnk is added to crude cell extracts.

Figure 3.2

Expression of a 33 kd polypeptide by recombinant T4 and λ phages. Autoradiographic analysis of ^{35}S labelled polypeptides following infection of UV irradiated cells (M159) with T4 or λ phage. Polypeptides were separated by electrophoresis through a 15% linear SDS polyacrylamide gel. Infections with T4 were labelled from 3-7 min and those with λ derivatives from 3-13 min or 20-30 min. The positions of standard marker proteins are indicated: phosphorylase b, (94 kd); bovine serum albumin (67 kd); ovalbumin (43 kd); carbonic anhydrase (30 kd): soybean trypsin inhibitor (20.1 kd). The tracks are: (a) uninfected cells (M159); (b)-(g) a series of T4 am phage derived from a cross of T4amN122 and T4pseT2; (h) T4D (wild-type); (i) T4pseT2 (a pseT$^-$ strain); (j) T4amN122; (k) λNM1108 labelled 3-13 min; (l) λNM1108 labelled 20-30 min; (m) λNM1108 deleted for fragment 21 labelled 3-13 min; (n) λNM1108 deleted for fragment 21 labelled 20-30 min. The position of the 33 kd polypeptide and of the 20 kd polypeptide with varying mobility are indicated.

The presence of kinase activity on infection with the T4 recombinant phages was compared with the pattern of polypeptides obtained on infection of UV irradiated cells (Figure 3.2). A T4 recombinant was found (Figure 3.2, track (b)) which expressed a 33 kd polypeptide and a 20 kd polypeptide with mobility identical to that of the corresponding T4 pseT2 (pseT$^-$) polypeptide. Since this recombinant T4 phage induced kinase activity on infection of non-irradiated cells (Figure 3.3, track (e)), kinase activity in T4 infections appeared to correlate exclusively with expression of the 33 kd polypeptide.

3.1.b. Expression of a 33 kd polypeptide by a λ recombinant including T4 EcoRI Fragment 21

The source of fragment 21 DNA was λNM1210, the recombinant derivative of λNM607 isolated by Mileham et al. (1980). This recombinant includes fragment 21 within the coding sequence of the cI gene. DNA inserted into the cI gene can be transcribed leftward from p_{RE} like the cI gene itself but there is no published evidence to suggest that the anti-sense strand of the cI gene is transcribed rightward from a λ promoter. Alternative λ vectors offer potential for transcription irrespective of the orientation of the inserted DNA. EcoRI fragment 21 was therefore transferred from λNM1210 to the central region of an integration-proficient λ vector, λNM459, resulting

Figure 3.3

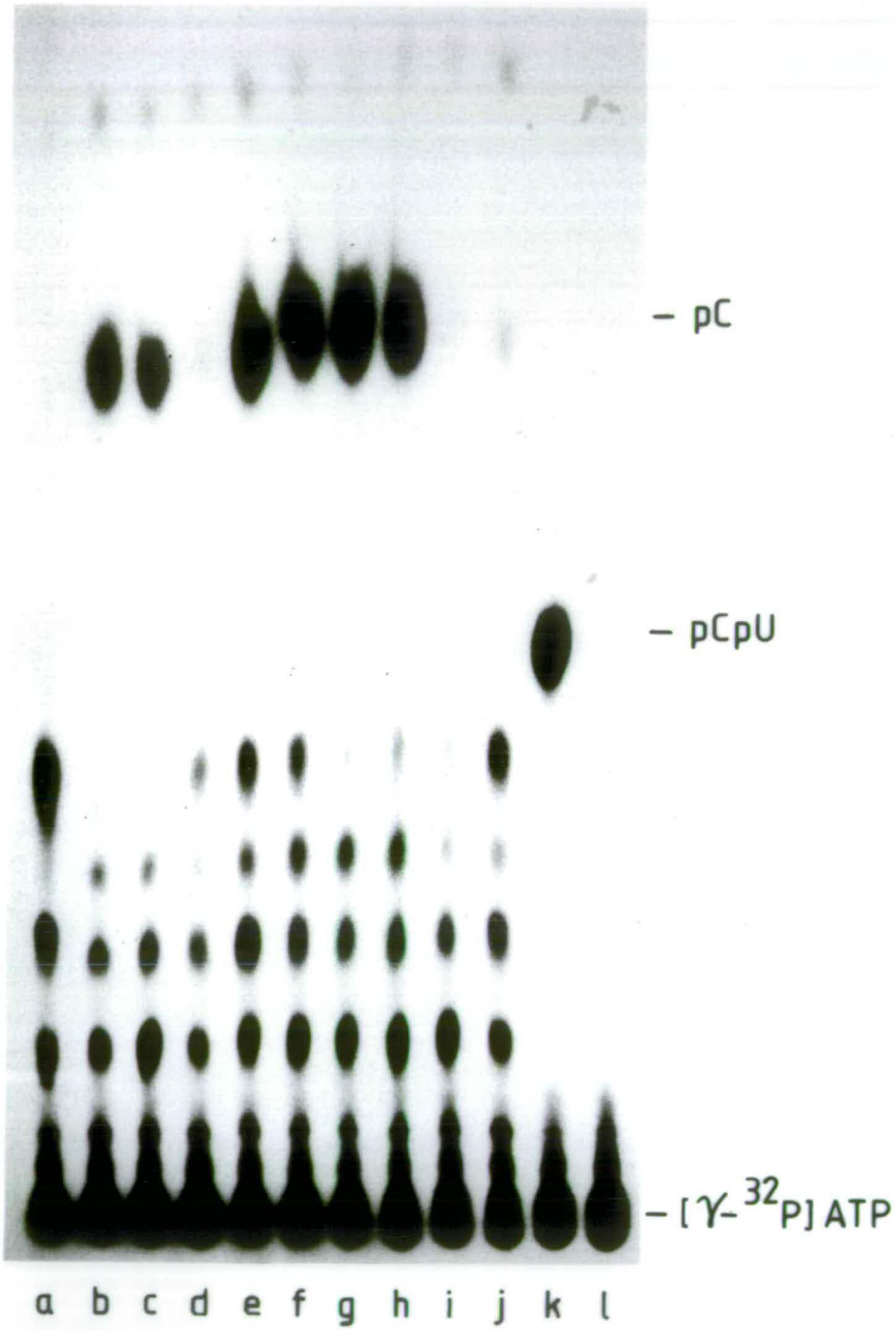


Figure 3.3

Assay for kinase activity in crude extracts of non-irradiated cells (ED8689) infected with T4 phage. The assay involves the conversion of the substrate CpU to the dinucleotide pCpU by transfer of a labelled phosphate from [γ-^{32}P]ATP. A nuclease in crude extracts degrades pCpU to pC (see section 3.1.a). Products of the reaction were separated by ionophoresis on AE81 cellulose paper at pH 3.5. Tracks are: (a) uninfected cells (ED8689); (b) T4D (wild-type); (c) T4amN122; (d) T4pseT2 (a pseT$^-$ strain); (e)-(j) a series of T4am phage derived from a cross of T4amN122 and T4pseT2, these correspond to tracks (b), (g), (c), (h), (d) and (i) respectively of Figure 3.2. In Figure 3.2, kinase activity in T4 phage infections seems to correlate with the presence of a 33 kd polypeptide and the higher mobility form of the 20 kd polypeptide (see section 3.1.a). However, the T4 recombinant used in track (b) of Figure 3.2 expresses a 33 kd polypeptide and the lower mobility form of the 20 kd polypeptide but does induce kinase activity (track (e) of this figure). Track (k) is an assay with purified T4 encoded pnk; track (l) is an assay with distilled water. The positions of [γ-^{32}P]ATP and of the labelled products of the kinase reaction, pCpU or pC are indicated.

in the recombinant phage λNM1104 (Figure 3.4).

UV irradiated bacteria (M159) were infected with recombinant λ phage and [^{35}S]methionine labelled polypeptides were analysed by SDS polyacrylamide gel electrophoresis (Figure 3.2). λNM1108, the Wam Eam Sam derivative of λNM1104, but not a derivative of λNM1108 deleted *in vitro* for fragment 21, expressed a 33 kd polypeptide. This polypeptide migrated in a position corresponding to the 33 kd pseT polypeptide expressed by T4D (wild-type) and T4amN122 but absent in infections with the pseT$^-$ phage, T4 pseT2. This would be consistent with the location of the pseT gene in fragment 21, although on some gels the polypeptide expressed by the recombinant λNM1108 appeared to have a very slightly faster mobility than the T4 pseT polypeptide (Figure 3.2), raising the possibility that λNM1108 might encode a truncated polypeptide.

3.1.c. Orientation and location of the ORF encoding the 33 kd polypeptide

The orientation of fragment 21 in λNM1104 and λNM1108 was deduced from the position of the asymmetrically located SmaI site (Figure 3.1). A recombinant, λCM1 (Figure 3.4) which carried fragment 21 in the opposite orientation to that in λNM1104 (and λNM1108) was isolated by ligation of fragment 21 purified from an EcoRI digest of λNM1210 to EcoRI cut λNM459. The orientation of fragment 21 in λNM1104/λNM1108

Figure 3.4

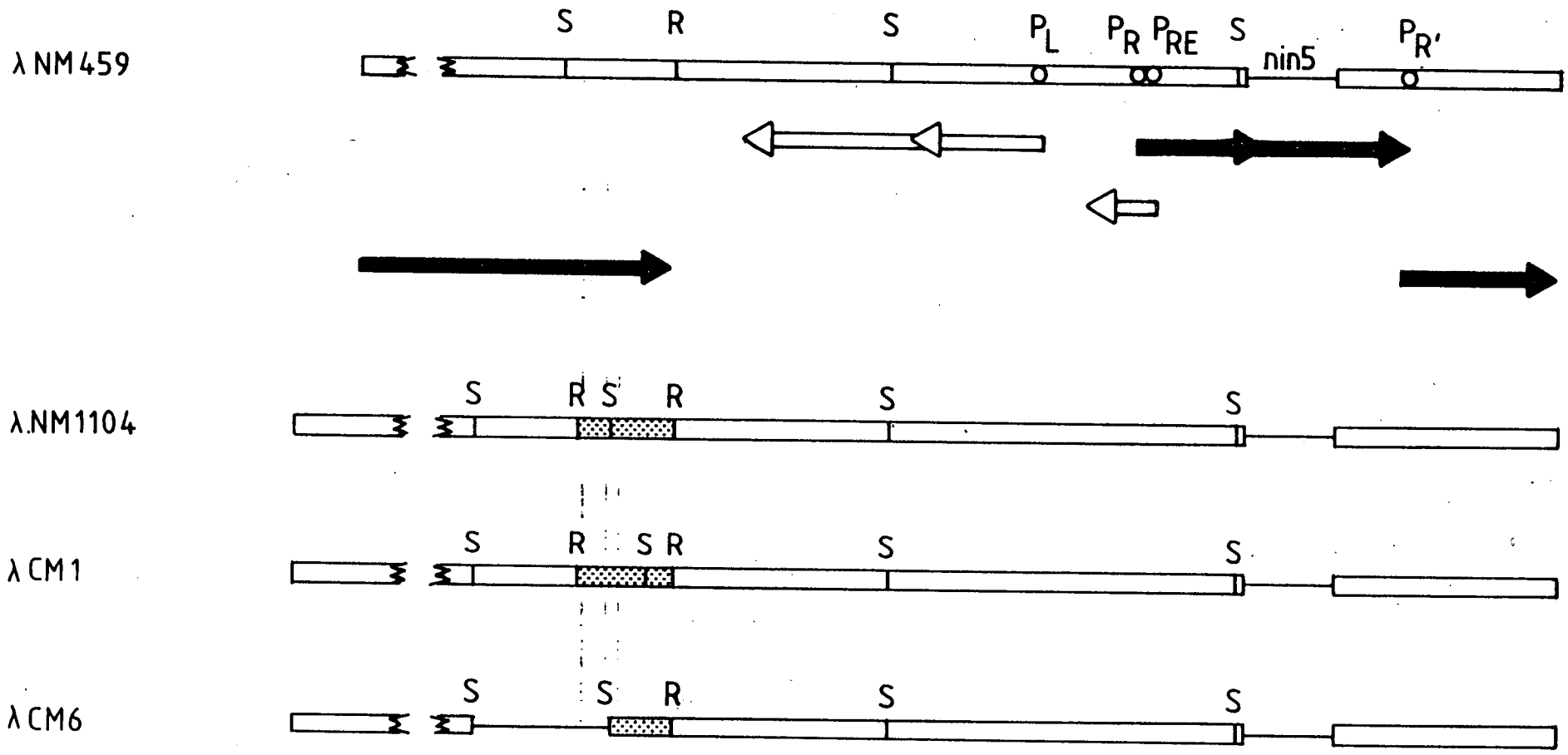
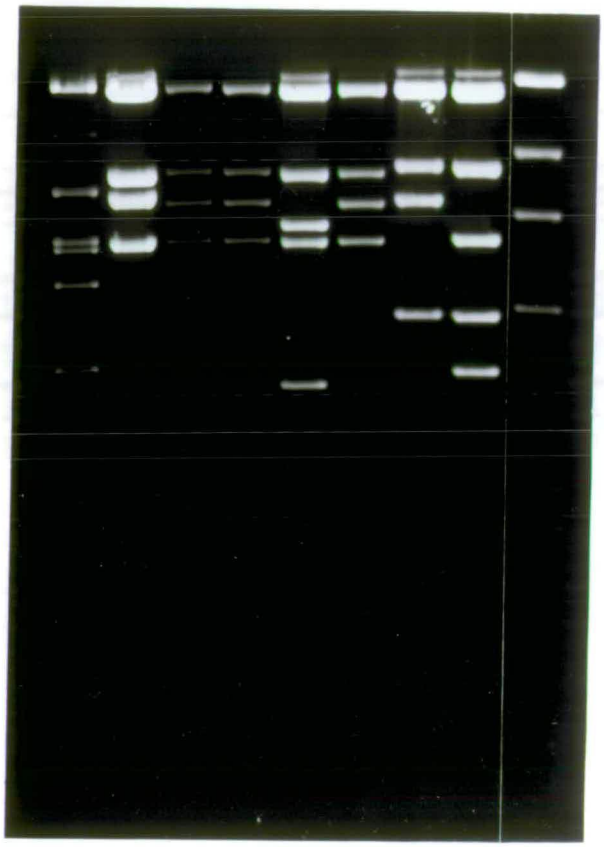

λ NM459
S
R
S
P_L
P_R
P_{RE}
S
nin5
$P_{R'}$
λ NM1104
S
R
S
R
S
S
λ CM1
S
R
S
R
S
S
λ CM6
S
S
R
S
S

Figure 3.4

Transcription patterns of vector NM459. The vector has a single EcoRI site (R) to the left of att. The genome circularises on infection and fragments inserted at the EcoRI site may be transcribed early from the leftward λ promoter p_L and late from the rightward promoter $p_{R'}$. Transcription from $p_{R'}$ requires activation by gpQ, a product of early rightward transcription, and must traverse more than 20 kb of the intervening λ genome before reaching the insert. λNM1104 and λCM1 are derivatives including T4 fragment 21 which encodes a 33 kd polypeptide. The respective orientations of the T4 fragments in λNM1104 and λCM1 were determined from SmaI digests (S indicates a SmaI site). λCM6 is like λCM1 but is deleted for the DNA between the two leftmost SmaI sites; this phage no longer expresses the 33 kd polypeptide. The open arrows show major leftward transcripts, the solid arrows the major rightward transcripts. T4 DNA is shaded.

and λCM1 resulted in fragment sizes of either 2.63 and 6.98 (Figure 3.5 tracks (c) and (d)) or 3.30 and 6.31 (Figure 3.5 track (e)) respectively for SmaI digests of the phage DNAs.

In λ, leftward transcription is mediated early from the powerful p_L promoter, and rightward transcription is dependent on the subsequent activation of the λ late promoter $p_{R'}$. The direction of transcription of the ORF in λNM1104 or λCM1, which was required for expression of the 33 kd polypeptide was deduced from the timing of expression of the polypeptide following infection of UV irradiated cells (M159) with those phages. Polypeptides were separated on 15% linear SDS polyacrylamide gels. For λNM1104, little expression of the polypeptide was detected when labelling was from 3-13 min after infection, whereas much more was detected when labelling was from 20-30 min after infection (Figure 3.6). This implied that transcription of the coding sequence required or was enhanced by activation of the late λ promoter $p_{R'}$. λCM1 expressed the 33 kd polypeptide more efficiently during the 3-13 min period following infection (Figure 3.6) indicating that transcription of the ORF was mediated by the λ early promoter p_L. These data identify the orientation of the sequence encoding the 33 kd polypeptide in relation to the λ promoters, and this can be related to the T4 map by the position of the



a b c d e f g h i

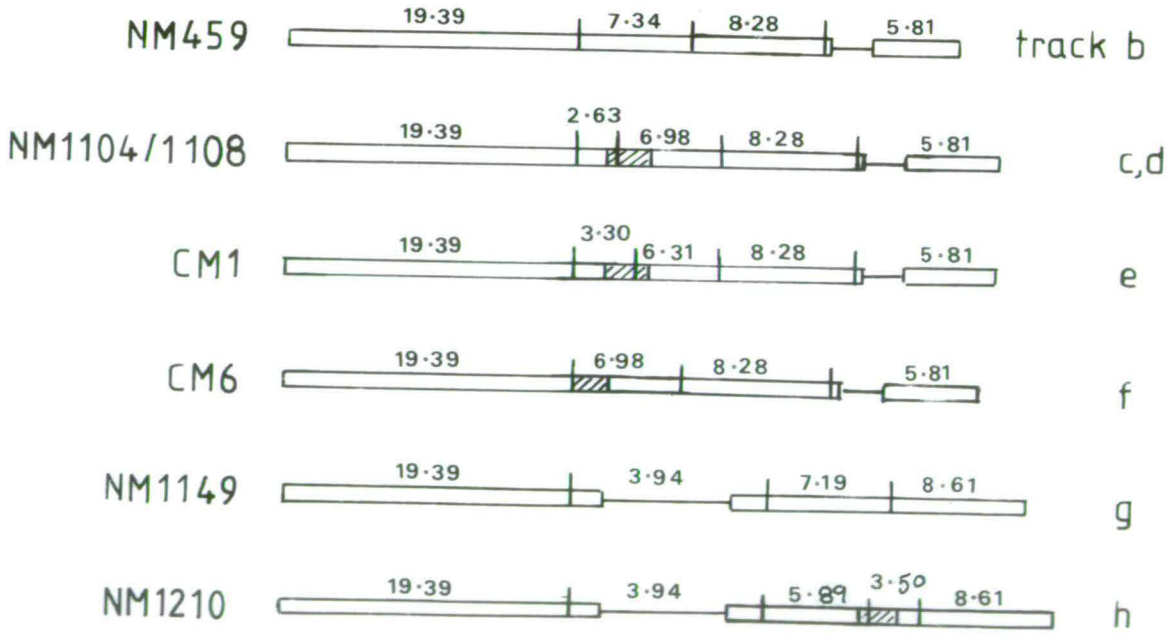


Figure 3.5

Figure 3.5

The orientation of fragment 21 in λ recombinants. SmaI digestion of λ recombinant DNAs including fragment 21. Fragments were separated by electrophoresis through a 0.8% agarose gel. Tracks are: (a) EcoRI digested λcI857 giving marker fragment sizes (kb) of 21.7, 7.52, 5.83, 5.64, 4.85 and 3.48; (b) SmaI digested λNM459; (c) SmaI digested λNM1104; (d) SmaI digested λNM1108; (e) SmaI digested λCM1; (f) SmaI digested λCM6; (g) SmaI digested λNM1149; (h) SmaI digested NM1210*; (i) HindIII digested λcI857 giving marker fragment sizes (kb) of 23.6, 9.64, 6.64, 4.34, 2.26, and 1.98. The origin of SmaI restriction fragments is shown in the accompanying diagram.

* The fragment calculated as 3.50 kb in a SmaI digest of λNM1210 migrated faster than expected. However the orientation of fragment 21 was confirmed from a HindIII - SmaI double digest cutting at the HindIII site 240 bp to the left of the insert yields a 1.0 kb HindIII - SmaI fragment rather than a 1.7 kb fragment.

asymmetrically located SmaI site in fragment 21 (Figure 3.1). The coding sequence is transcribed anti-clockwise with respect to the T4 genome (i.e. left to right in Figure 3.1), as would be expected for a gene which is expressed early in the T4 life cycle.

Some idea of the location of the sequence encoding the 33 kd polypeptide was gained from λCM6, a recombinant which was deleted for the smaller of the two SmaI-EcoRI fragments derived from fragment 21 (Figure 3.4). To obtain this recombinant, SmaI cut fragment 21 was incubated in a mixed ligation with SmaI cut λNM1112 (to provide a SmaI left arm) and EcoRI cut λNM1222 (to provide an EcoRI right arm). The right arm of NM1112 and the left arm of λNM1222 carried amber mutations so that recombinant phage including these arms, or religated parental phage, would only grow on a host suppressor strain. Recombinants were therefore selected on a suppressor-free strain (NM594) and those including the larger SmaI-EcoRI fragment were identified by digestion with SmaI (e.g. λCM6 in Figure 3.5). On infection of the UV irradiated cells, λCM6 failed to express a 33 kd polypeptide (Figure 3.6) either early or late in infection suggesting that the SmaI site is within the gene coding for the 33 kd polypeptide.

Figure 3.6

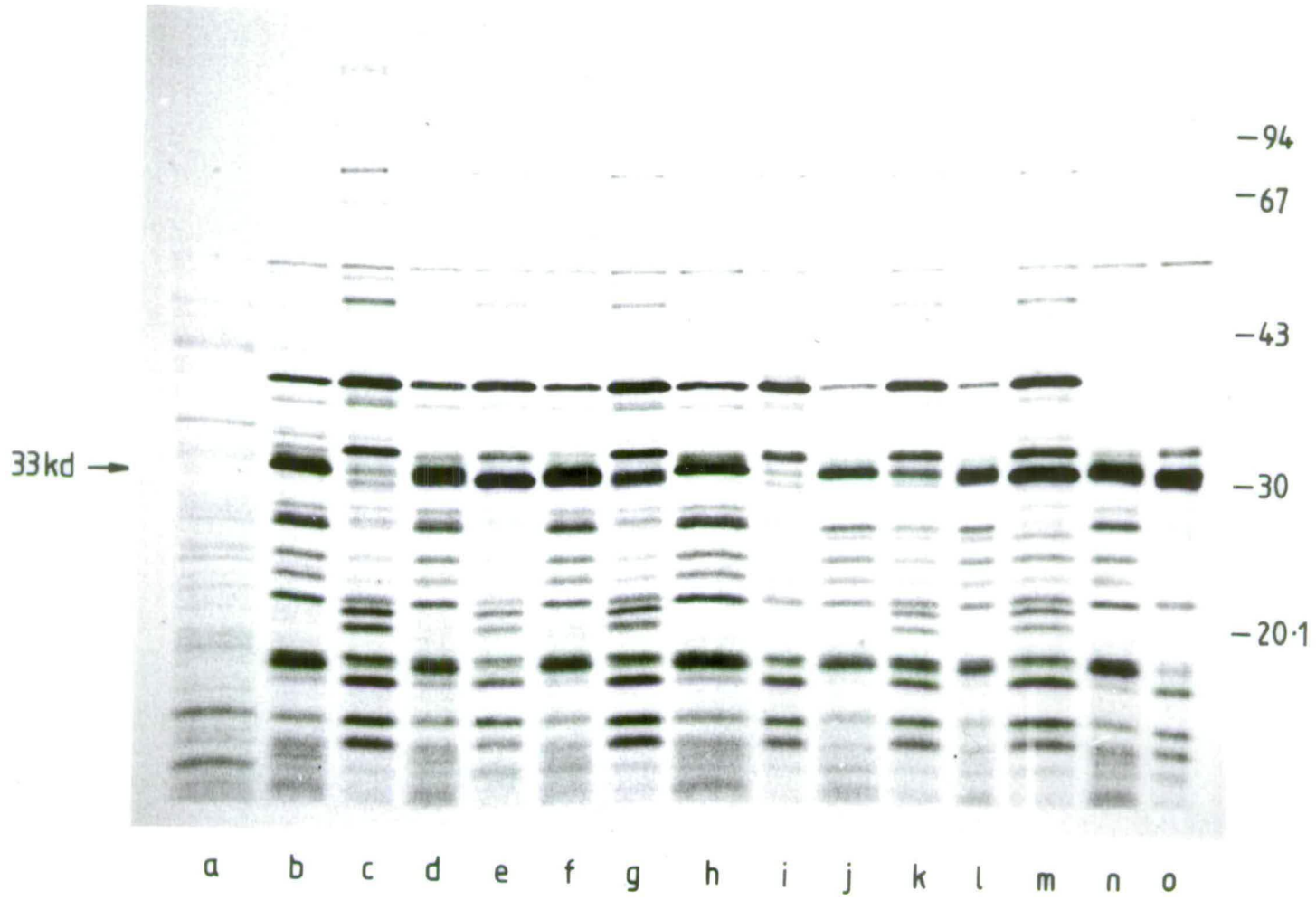


Figure 3.6

The timing of expression of the 33 kd polypeptide by λ recombinants including fragment 21 or fragment 21^{+}. Autoradiographic analysis of ^{35}S labelled polypeptides following infection of UV irradiated cells (M159). Polypeptides were separated by electrophoresis through a 15% linear SDS polyacrylamide gel. Tracks are (a) uninfected cells (M159); (b) λNM459 vectors labelled 3-13'; (c) λNM459 vector labelled 20-30'; (d) λNM1104 labelled 3-13'; (e) λNM1104 labelled 20-30'; (f) λCM1 labelled 3-13'; (g) λCM1 labelled 20-30'; (h) λCM6 labelled 3-13'; (i) λCM6 labelled 20-30'; (j) λNM1149 vector labelled 3-13'; (k) λNM1149 vector labelled 20-30'; (l) λCM8 labelled 3-13'; (m) λCM8 labelled 20-30'; (n) λCM21 labelled 3-13 min; (o) λCM21 labelled 20-30'. The positions of standard marker proteins are indicated (see figure 3.2). The position of the 33 kd pnk polypeptide is also shown.

3.1.d. λ clones including fragment 21 do not induce kinase activity

Crude cell extracts from non-irradiated cells (ED8689) infected with λ phage at high multiplicity were assayed for kinase activity by the ability to transfer ^{32}P from [γ-^{32}P]ATP to the dinucleoside monophosphate CpU (Figure 3.7). No activity was detected in extracts from suppressor-free cells (ED8689) infected with λNM1108 (Figure 3.7 track (1)), a phage that on infection of UV irradiated cells produced large amounts of a 33 kd polypeptide (Figure 3.2). The polypeptide encoded by λ recombinants including fragment 21 was therefore defective as the result of either a mutation or incorrect processing, or alternatively was not the product of the pseT gene.

Repeated attempts were made to clone a T4 fragment extending from the SmaI site in fragment 21 to a downstream BglII or PstI site (see Figure 3.1) to determine whether DNA downstream of fragment 21 encoded an alternative 33 kd polypeptide. These attempts were unsuccessful perhaps because the recombinants included DNA encoding functions that were deleterious to phage or host propagation. The 1.3 kb EcoRI fragment upstream of fragment 21 (Figure 3.1) had not been cloned in λ or plasmid vectors (Mileham et al., 1980). In addition T4pseTΔ1 had been reported to be alc$^-$ (Snyder et al., 1976) which indicated that the alc/unf gene might be located

Figure 3.7

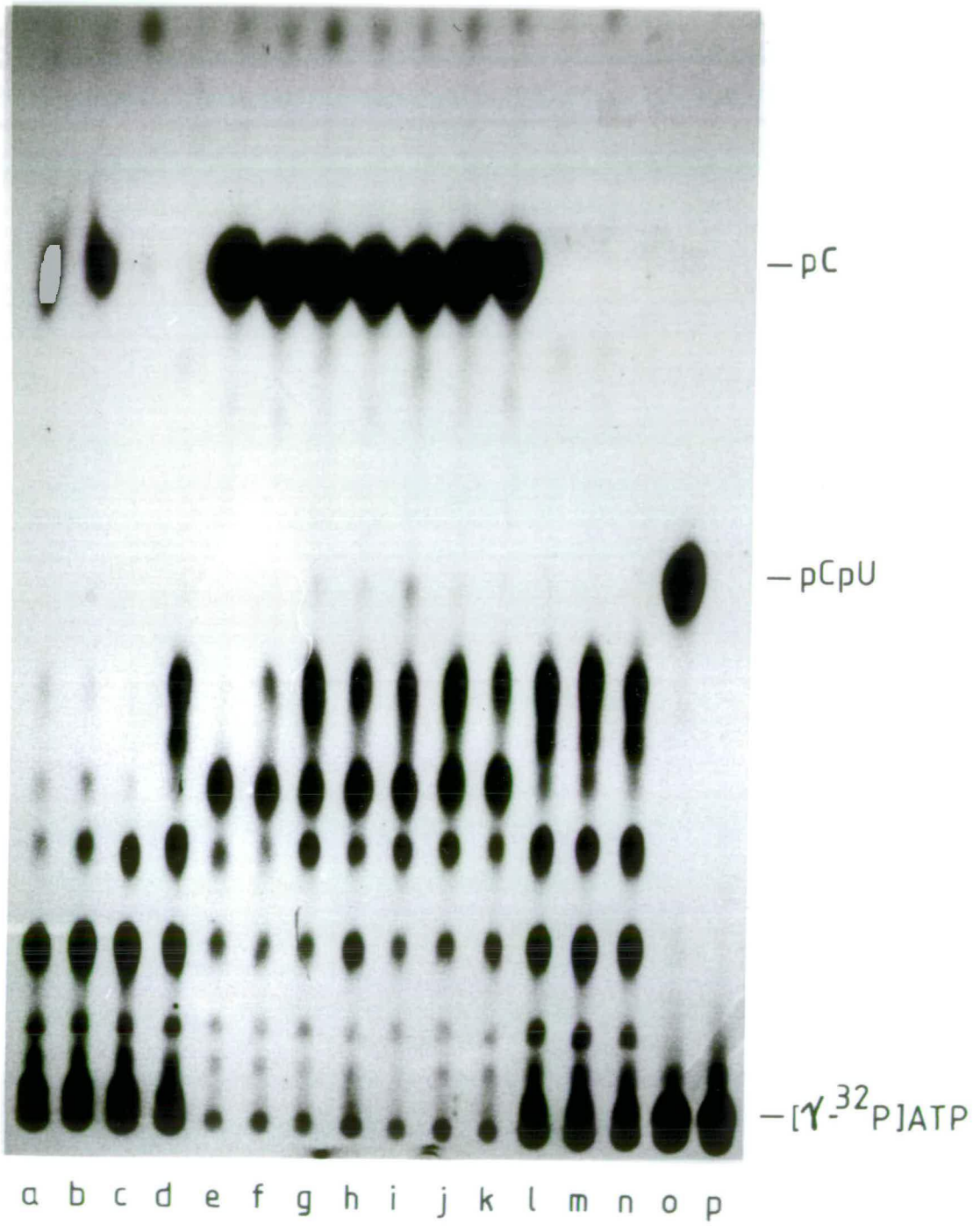


Figure 3.7

Assay for kinase activity in crude cells extracts (ED8689) after infection with λ recombinants including fragment 21 derived from λNM1210 or fragment 21^{+} derived directly from T4 DNA. Radiolabelled products of the kinase reaction were separated by ionophoresis on AE81 cellulose paper at pH 3.5. Tracks are: (a) T4D (wild-type); (b) T4amN122; (c) T4pseT2 (a pseT^{-} strain); (d) NM1149 vector; (e)-(k) seven different lysates of λCM8 which includes EcoRI fragment 21^{+} derived directly from T4 DNA; (l) λNM1108; (m) λNM1108 deleted for EcoRI fragment 21; (n) uninfected cells (ED8689); (o) purified T4 encoded pnk; (p) distilled water. The positions of [γ-^{32}P]ATP and of the labelled products of the kinase reaction, pCpU or pC are indicated.

close to fragment 21 and may even extend into this fragment. Sequences upstream of fragment 21 were therefore considered unlikely to encode a 33 kd pseT polypeptide.

3.1.e. The DNA sequence of EcoRI fragments 21 and 46

All of the DNA sequence information was derived from sequences cloned in M13 vectors using the dideoxy chain termination method (section 2.2.q.). T4 EcoRI fragments 21 and 46 were purified from EcoRI digests of λNM1210 and λNM1207 respectively. Fragment 21 was ligated to EcoRI cut M13mp11 and fragment 46 was ligated to EcoRI cut M13mp8. Both strands of fragment 46, and the extreme ends of fragment 21 were sequenced.

To obtain sequence for the rest of fragment 21, the fragment was sonicated to generate random fragments 300-600 base pairs in length. After repairing the ends with polymerase, these fragments were ligated to SmaI cut M13mp8 vector. Initially about 60 of these recombinants were sequenced, and the random stretches of fragment 21 DNA sequence were compiled and ordered on a VAX-VMS computer using the programmes of Staden (1982). By this method, some parts of the sequence were covered more often than others and in some areas, sequence could not be obtained on both strands. To complete these gaps, clones of known sequence were used as single-stranded

```
1321  GGTTATCAAATCGTAGTCGTTTCAGGTCGTGAAAGTGGAACTAAAGAAGACCCAACGAAA
      GlyTyrGlnIleValValValSerGlyArgGluSerGlyThrLysGluAspProThrLys

1381  TATTATCGTATGACCCGTAAATGGGTTGAGGACATTGCTGGCGTTCCATTAGTTATGCAA
      TyrTyrArgMetThrArgLysTrpValGluAspIleAlaGlyValProLeuValMetGln

1441  TGTCAGCGCGAACAAGGCGATACCCGTAAAGACGATGTAGTTAAAGAAGAAATTTTCTGG
      CysGlnArgGluGlnGlyAspThrArgLysAspAspValValLysGluGluIlePheTrp

1501  AAACACATTGCACCGCATTTTGACGTGAAATTAGCTATTGATGACCGAACTCAAGTAGTT
      LysHisIleAlaProHisPheAspValLysLeuAlaIleAspAspArgThrGlnValVal

1561  GAAATGTGGCGTCGTATCGGTGTTGAATGCTGGCAAGTCGCTTCGGGAGATTTTTAATGG
      GluMetTrpArgArgIleGlyValGluCysTrpGlnValAlaSerGlyAspPheEnd

1621  CTTGGCACCATGAAACTTGGGCTATTGTTATTGTAAATAGCGGTTTAGTTGGTACTAGTA

1681  ATGGGCAATTTTGTGTATTTACTAGTGAAAATAGAGCATGGGAGGAATGTCTTAAATTAA

1741  GAGAAAAGAATCCCGATGTTGAACTAGTAGTAAAGAAAACTAAACTGCCTTTACCATGGA

1801  AAACTTATGAATAACCTAGAAAAGATTTATCGTCTTTGTGATAAAATTGAAAAAGAAAAG

1861  AAATATCTATTTTGTCTATGGCCTATTGTTGACGGAAGAGTAGGCCTAGATGTTCTTGAT

1921  TATGAAACAGAAGACAGAGTAGATGGTTCAACTTTTGATAATGCGTTGGATGTTATTGAT

1981  TGGCTTGAAGAAAATTATGTGAGGTAAATATGTTTCCGACTTACTCTAAAATCGTAGAAG

2041  TAGTGTTTAGCCAAATTATCGCTAATAATATGTTTGAAAAACTTGATAACGCAGCTGAGC

2101  TTCGAATCCATGCTCAAGTGACTCATGTATTGAACACTTTGCTTCCAGACCAGGTGGATT

2161  CTATTGCCATTACACTGTATCCAGGTTCCGCGCATATCATTGTCGTATTTGGTCTTGATG
                                                        EcoRI
2221  CTGAGCTTGTTATCAAAGGCGATATTCGCTTTGAATCTCAAACTGCGGAATTCAAAGCGA

2281  TTTAATGGTTTACTTTACGGTAGAGCTATAATATCACAACTCTACCAAAACAAATGAGGA
                                                    EcoRI
2341  AAACAAAATGTTGCTAAGTGAAAAACCGATTACTGTTAAAGGAATTC
```

Figure 3.8

```
      EcoRI
1     GAATTCAGTATAATTATATTGATGCGATGAATAATAAAAATCGTGAGGCAATTGCTGCTA

61    TTGAGCGTGAAAATGAAAAACTGCGCAAAGATGCAAAGAAGGCGGATGTGGTGGCTCATA

121   AGCCAGGATTGGTTGAAAAACAAATCAACAACTCCTTCAACAAGTTCGCAGAAGACATCC

181   AGGACCTTTCTAAATGATTAAACTATCAGCAGTAATATTATCTATTGGTCTTCTAGTTGG
                        XbaI
241   TTGTTCGACAAAGCCTCTAGAAGTAAAGAAAGAAACAGTTCATCCTAATTGGCCTGTGCA

301   AATAAAGTCATATGATGAAGCTAAACTATCTTGGCAAGTTAAAGTTATTGATGGTAAAGC

361   CTGGGTCGGTATGCCATTTGAAGATTCTCAGGAATTTCGTATTTGGCTTAATGATGTAAA

421   ACGATATGTACATGACCAGAAAACTATGAXATGTTATTATCGTCAAGAGCTAAAAGAGGA

481   TAAATGTAAATGATTTCATGGCATCAATTTGAACATCTCAAAGGATTGATTTATGAATCC

541   GAGATGGCTGCAATGATTTATGGACGCCAGATTCAGCGGTTAGAATCTTTACCTCCAACT

601   AATGATGTTTTATTAGCTCAATCACGTGCTAATCTCAAAAATGAATATCAAAATAAGTGG

661   GGTAAAGCATCTAAAGACCTACATGATTATATTCAATCATTAGTTGAGAAAAAATAAATG
                                                              Met

721   AAAAAGATTATTTTGACTATTGGCTGTTCTGGTAAGAGTACTTGGGCTCGTGAATTTATT
      LysLysIleIleLeuThrIleGlyCysSerGlyLysSerThrTrpAlaArgGluPheIle
                  SmaI
781   GCTAAGAATCCCGGGTTTTATAATATCAATCGTGATGACTATCGCCAATCTATTATGGCG
      AlaLysAsnProGlyPheTyrAsnIleAsnArgAspAspTyrArgGlnSerIleMetAla

841   CATGAAGAACGCGATGAGTACAAGTATACCAAAAAGAAAGAAGGTATCGTAACTGGTATG
      HisGluGluArgAspGluTyrLysTyrThrLysLysLysGluGlyIleValThrGlyMet

901   CAGTTTGATACAGCTAAAAGTATTCTGTACGGTGGCGATTCTGTTAAGGGAGTAATCATT
      GlnPheAspThrAlaLysSerIleLeuTyrGlyGlyAspSerValLysGlyValIleIle

961   TCAGATACTAACCTGAATCCTGAACGTCGCCTAGCATGGGAAACTTTTGCCAAAGAATAC
      SerAspThrAsnLeuAsnProGluArgArgLeuAlaTrpGluThrPheAlaLysGluTyr

1021  GGCTGGAAAGTTGAACATAAAGTGTTTGATGTTCCTTGGACTGAATTGGTTAAACGTAAC
      GlyTrpLysValGluHisLysValPheAspValProTrpThrGluLeuValLysArgAsn

1081  TCAAAACGCGGAACTAAAGCAGTACCAATTGATGTTTTACGTTCAATGTATAAAAGCATG
      SerLysArgGlyThrLysAlaValProIleAspValLeuArgSerMetTyrLysSerMet

1141  CGAGAGTATCTCGGTCTTCCAGTATATAATGGGACTCCTGGTAAACCAAAAGCAGTTATT
      ArgGluTyrLeuGlyLeuProValTyrAsnGlyThrProGlyLysProLysAlaValIle

1201  TTTGATGTTGATGGTACACTAGCTAAAATGAATGGTCGTGGTCCTTATGACCTTGAAAAA
      PheAspValAspGlyThrLeuAlaLysMetAsnGlyArgGlyProTyrAspLeuGluLys

1261  TGCGATACCGATGTTATCAATCCTATGGTTGTTGAACTGTCTAAGATGTATGCTCTTATG
      CysAspThrAspValIleAsnProMetValValGluLeuSerLysMetTyrAlaLeuMet
```

Figure 3.8

The DNA sequence of the contiguous EcoRI fragments 21 and 46 with the amino acid sequence of the 900 base ORF encoding the 33 kd polypeptide. EcoRI fragments 21 and 46 were derived from λNM1210 and λNM1207 respectively (Mileham et al., 1980). Restriction sites for EcoRI, SmaI and XbaI are shown above the DNA sequence and the translation of the ORF is shown below.

probes to screen the library of random fragments in M13 to find the required complementary sequences. One section (between bases 302 and 464 in Figure 3.8) which was not represented at all in the library on one strand was sequenced from the M13mp11 recombinant including the intact fragment 21 using specific synthetic oligonucleotides

(5'-GTTCGACAAAGCCTA and

5'-AAACAGTTCATCCTAAT)

instead of the usual universal primer.

The orientation of fragment 21 in relation to the physical and genetic map of T4 was indicated by the unique SmaI site (see Figure 3.1). The orientation of fragment 46 was determined from the sequence of an AluI fragment derived from an AluI digest of the T4 8.2 kb BglII fragment (see Figure 3.1) which includes EcoRI fragment 21 and 46. Fragments from the AluI digest were ligated to SmaI-cut M13mp18 vector DNA. An M13-AluI fragment recombinant which hybridised to both λNM1210 and λNM1207, i.e. overlapped fragments 21 and 46 was chosen for sequencing the junction between the two fragments.

The sequence revealed only one large ORF of 900 bases (Figure 3.8) which was contained entirely within fragment 21. Several smaller ORFs, none larger than 320 bases occurred in this region (Figure 3.9), however, no large ORF extended over fragments 21 and 46 and into fragment 33 since stop codons

Figure 3.9

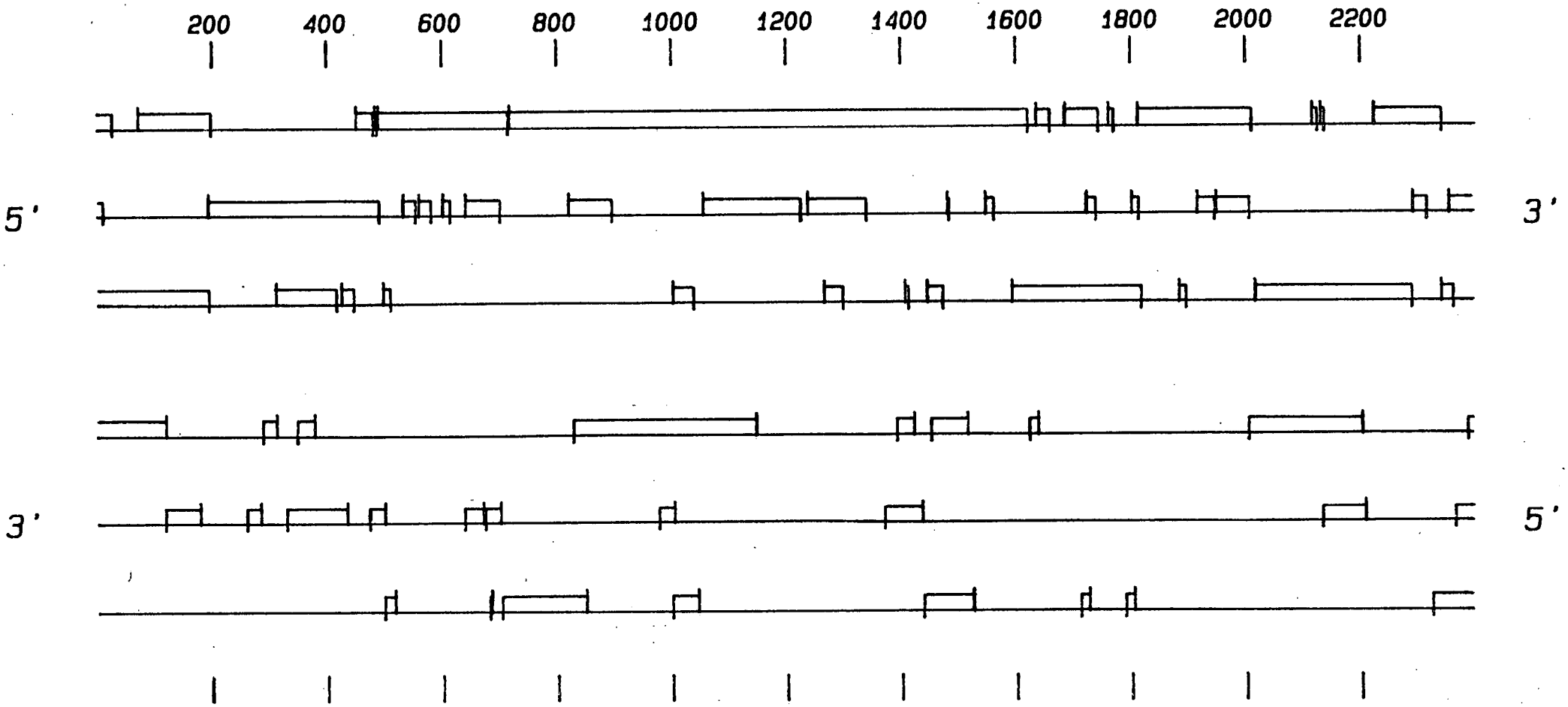

200
400
600
800
1000
1200
1400
1600
1800
2000
2200
5'
3'
3'
5'
FRAMES of: KINCON.SEQ Check: 8556 from: 1 to: 2393

Figure 3.9

A map indicating sizes of the ORFs in all six reading frames of the contiguous sequence of EcoRI fragments 21 and 46.

blocked all frames in fragment 46.

The direction of transcription predicted for the 900 base ORF corresponded to that determined for expression of the 33 kd polypeptide in λNM1104. The proposed ATG initiation codon for the ORF was situated 75 bases upstream of the SmaI site in fragment 21. This was consistent with the loss of expression of the 33 kd polypeptide in the λ recombinant λCM6 which is deleted for T4 DNA upstream of this SmaI site (Figure 3.4).

The large ORF located entirely within fragment 21 therefore appeared to encode the 33 kd polypeptide expressed by λ recombinants e.g. λNM1104. Expression of the 33 kd polypeptide did not confer kinase activity on infections with these λ recombinants, therefore either the 33 kd polypeptide is a defective pnk polypeptide or the pseT polypeptide is encoded by an alternative ORF. Some support for the former interpretation was derived from the analysis of the polypeptides expressed by the T4pseT$^+$ strains (Figure 3.2); there was only a single polypeptide band in the 33 kd pseT polypeptide region which was missing in T4pseT2 (pseT$^-$) infections. One possibility for the expression of a defective pseT polypeptide was indicated by the DNA sequence. Lillehaug (1977) reported that the N-terminal amino acid of the pnk subunit was phenylalanine, however the first phenylalanine in the sequence of the 900

Figure 3.10

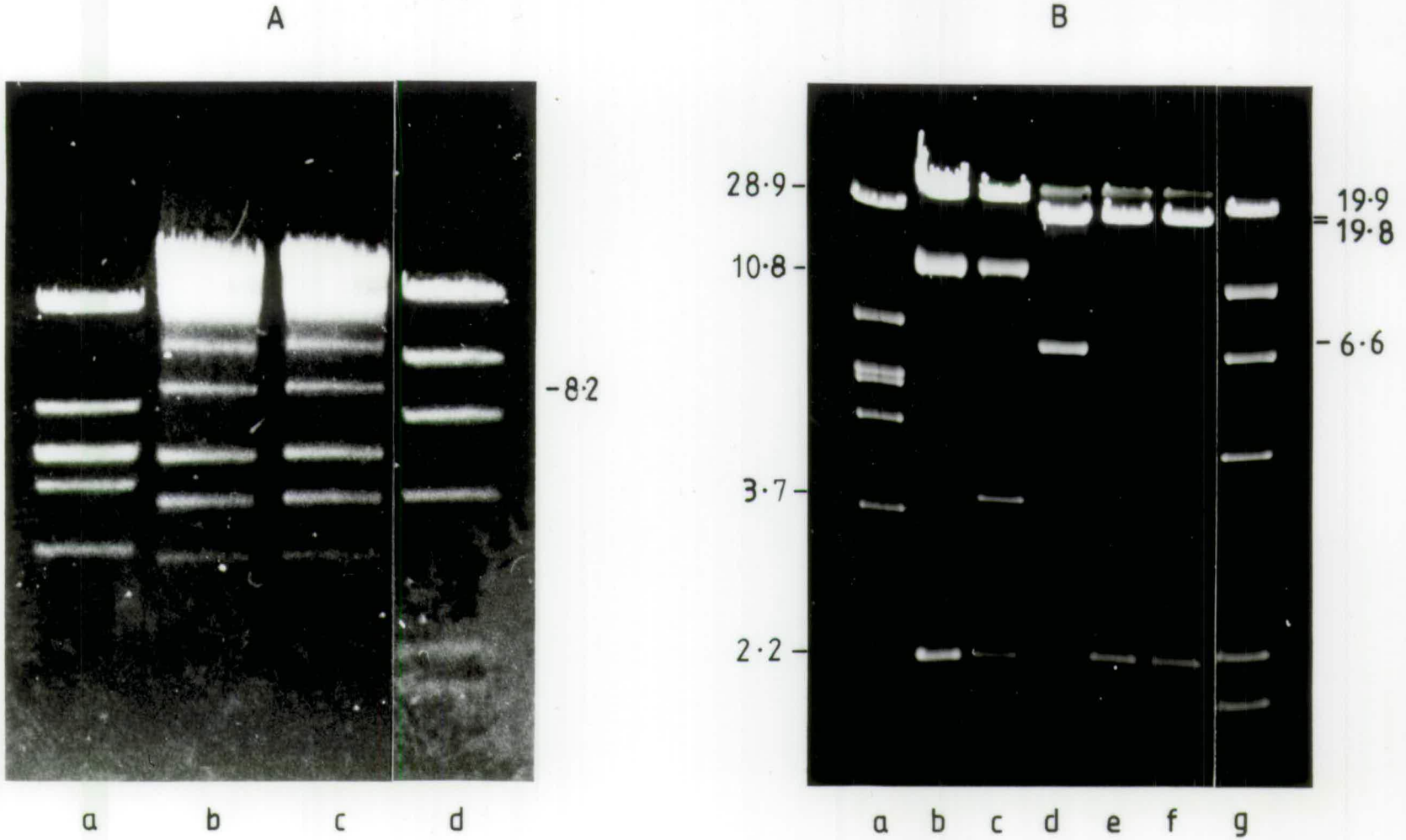


Figure 3.10

A. BglII digestion of T4 alc7 DNA. Fragments were separated by electrophoresis through a 0.8% agarose gel. Tracks are: (a) EcoRI digested λ cI857 DNA giving marker fragment sizes (kb) of: 21.7, 7.52, 5.83, 5.64, 4.85, and 3.48; (b) and (c) BglII digested T4 alc7 DNA; (d) HindIII digested λcI857 DNA giving marker fragment sizes (kb) of: 23.6, 9.64, 6.64, 4.34, 2.26 and 1.98. The position of the T4 8.2 kb BglII fragment, which includes T4 EcoRI fragments 21, 46 and 33 (Figure 3.1) is indicated.

B. EcoRI digestion of recombinants including an active pseT gene. Tracks are (a) EcoRI digested λcI857 DNA (see above); (b) EcoRI digested λCM8; (c) an EcoRI digest of a second recombinant derivative of λNM1149 which encodes an active pnk (Figure 3.16, track (o)) but also includes an unidentified EcoRI fragment of 3.7 kb; (d) EcoRI digested λNM1070; (e) EcoRI digested λCM21; (f) an EcoRI digest of a second recombinant derivative of λNM1070 which is like λCM21; (g) HindIII digested λcI857 DNA. The positions of the lacZ fragment of λNM1070 (6.6 kb) and of fragment 21 (2.2kb) are indicated.

base ORF within fragment 21 was at residue 20. This raised the possibility of a failure of post-translational processing. However, the pseT polypeptide expressed by T4pseT$^+$ phage, and the 33 kd polypeptide expressed by λNM1104 or λNM1108 have very similar mobilities, which did not support the idea of an T4 dependent removal of 20 amino acid residues from the inactive 33 kd polypeptide. Processing by the addition of a terminal phenylalanine remained a possibility. The alternative explanation for a defective 33 kd polypeptide, i.e. a mutation within the cloned ORF, was investigated by the isolation of further clones containing fragment 21 derived from a different DNA preparation.

3.1.f. Isolation of clones encoding an active pnk

The 8.2 kb BglII fragment (Figure 3.10A), which includes T4 EcoRI fragment 21, was purified from T4 alc7 DNA and digested with EcoRI. The resulting fragments were cloned in the immunity region of vector λNM1149 (this differs from λNM607 only in that it may be used for either EcoRI or HindIII fragments). Recombinants recovered as clear plaques on a hfl$^-$ host (NM514) were screened by hybridisation to a probe made from purified fragment 21 to identify those including an alternative T4 2.27 kb EcoRI fragment (fragment 21$^+$). Two recombinants including fragment 21$^+$ were recovered and high titres lysates were prepared (sections 2.2.f and 2.2.g). Crude

Figure 3.11

The orientation of fragment 21^+ in λ recombinants encoding an active pnk. SmaI digestion of λ recombinant DNAs including fragment 21^+. Fragments were separated by electrophoresis through a 0.8% agarose gel. Tracks are: (a) EcoRI digested λcI857 giving marker fragment sizes (kb) of 21.7, 7.52, 5.83, 5.64, 4.85 and 3.48; (b) SmaI digested λCM8*; (c) SmaI digested λ1149; (d) SmaI digested λCM21; (e) a SmaI digest of a second recombinant of λNM1070 including fragment 21^+ which is like λCM21; (f) SmaI digested λNM1070; (g) HindIII digested λcI857 giving marker fragment sizes (kb) of 23.6, 9.64, 6.64, 4.34. 2.26 and 1.98. The origin of SmaI restriction fragments is shown in the accompanying diagram.

* The fragment calculated as 3.50 kb in a SmaI digest of λcm8 migrates faster than expected. However the orientation of fragment 21 was confirmed from a HindIII - SmaI double digest (see legend to figure 3.5)

Figure 3.12

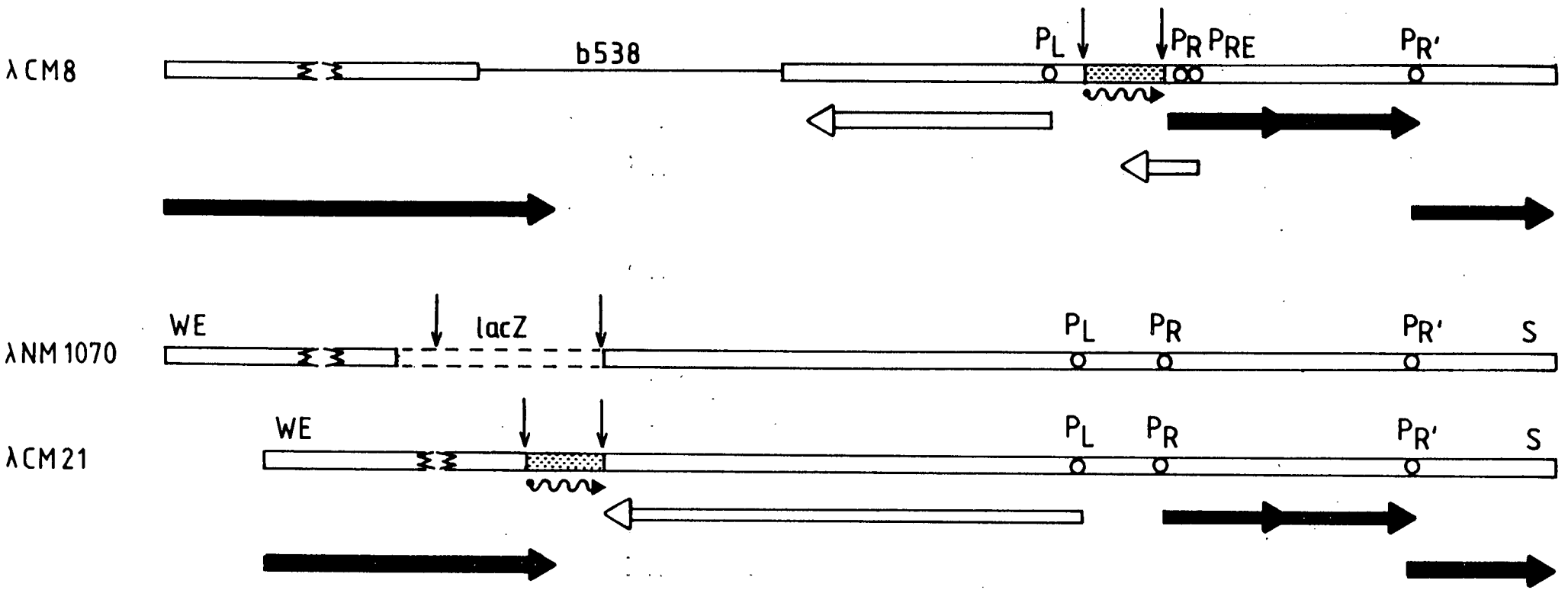

λCM8
b538
P_L
P_R
P_{RE}
$P_{R'}$
λNM1070
WE
lacZ
P_L
P_R
$P_{R'}$
S
λCM21
WE
P_L
P_R
$P_{R'}$
S

Figure 3.12

The origin and transcription of λCM21. The structure of λCM8, the first recombinant encoding an active pnk is shown. Fragment 21 is inserted in the orientation in which pseT cannot be transcribed from p_{RE}. Transcription of the pseT gene may depend on a promoter sequence within the insert, although transcription from a λ promoter has not been ruled out. In λCM21 T4 fragment 21 derived from λCM8 is inserted between the EcoRI site within the lacZ DNA and the EcoRI site of the λ DNA thereby replacing most of the lacZ DNA. Transcription of pseT will depend on, or be augmented by p_R, following amplification of the induced prophage. The amber mutations in genes E and W prevent the formation of virus particles, and that in S blocks cell lysis. Small arrows indicate EcoRI targets, the λ promoters and deletions are designated by their usual symbols. T4 fragment 21 is indicated as a shaded insert and the curly arrow beneath indicates the direction of transcription of pseT (the orientation of fragment 21 in λCM8 and λCM21 was determined from SmaI digests). Large open arrows show major leftward transcripts, and solid arrows the major rightward transcripts.

extracts from cells infected with each of these recombinants were assayed for kinase activity. In contrast to infections with λNM1108, phage including fragment 21^+ were found to induce kinase activity (Figure 3.7). SmaI digestion of DNA from one of these phage, λCM8 (Figure 3.11) gave the orientation of fragment 21^+ as that shown in Figure 3.12, i.e. transcription of the ORF encoding the 33 kd polypeptide would not be effected from p_{RE}. Since transcription of pseT from p_{RE} is ruled out, but expression of kinase activity was detected, fragment 21^+ may carry a functional promoter or alternatively transcription from p_R, or some unknown promoter is responsible for expression of the 33 kd polypeptide. No precedents for rightward transcription of DNA inserted in the cI region have been reported but it is possible that the b538 deletion, which removes 8.3 kb of the central region of the λNM1149 vector, removes terminators which normally prevent transcription from p_R, reaching DNA inserted in the immunity region.

3.1.g. The DNA sequence of the functional pnk gene

Fragment 21^+ was purified from an EcoRI digest of λCM8 DNA (Figure 3.10B). An AluI digest of this fragment was ligated to SmaI cut m13mp18. A sample of these M13 recombinants was used to obtain the DNA sequence of the ORF within fragment 21^+. Unfortunately the AluI fragment including the last 63 bases of the ORF was not recovered as an M13

Figure 3.13

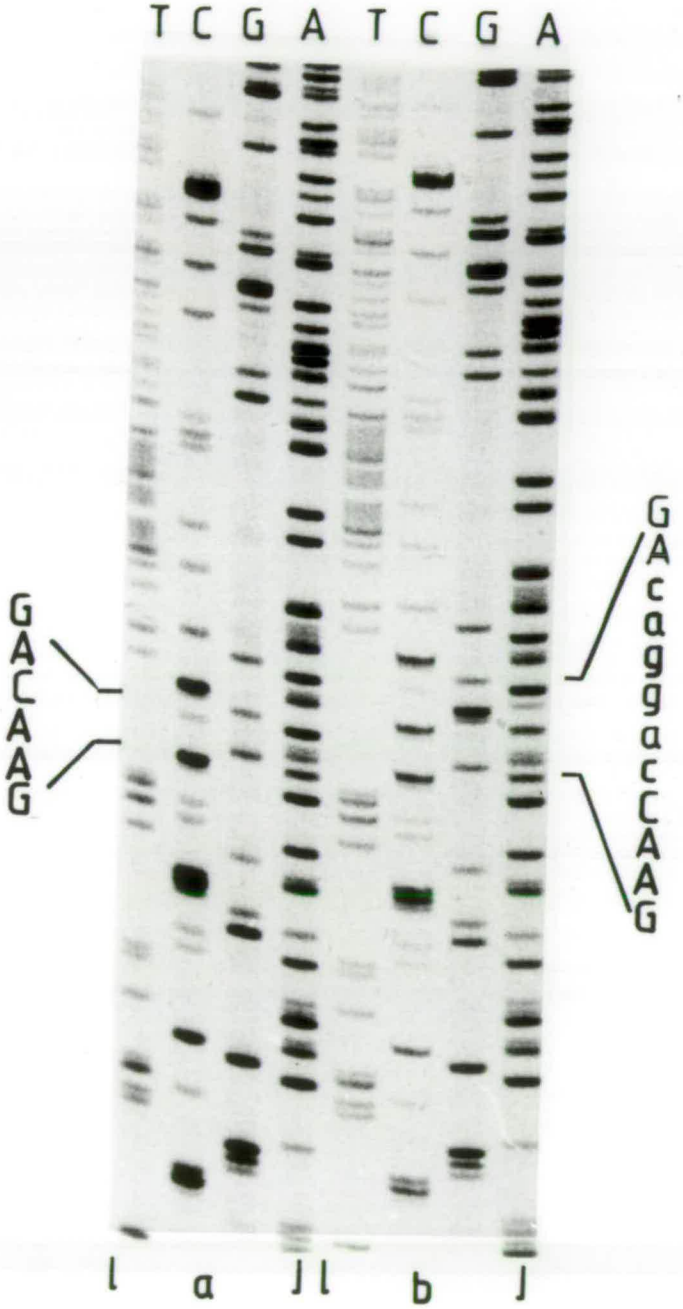


Figure 3.13

The DNA sequence of the region where a 6 base pair deletion had occurred in fragment 21 derived from λNM1210. AluI fragments cloned in M13mp18 were sequenced by the dideoxy chain termination method and separated by electrophoresis through a buffer gradient polyacrylamide gel. Tracks are: (a) an AluI fragment from a digest of fragment 21 derived from λNM1210; (b) an AluI fragment from a digest of fragment 21^{+} derived from λCM8. The order of tracks is given at the top of the figure. Sequence in the immediate region of the deletion is shown, lower case letters denote those bases

GTCCTG
CAGGAC

which are not present in fragment 21 derived from λNM1210.

recombinant. To obtain this sequence, it was therefore necessary to use a synthetic oligonucleotide (5'-TTCTGGAAACACATTGC) as a primer for an M13mp18 recombinant including the larger of the two SmaI-EcoRI fragments derived from fragment 21^+. Both of the SmaI-EcoRI fragments were obtained from a SmaI digest of fragment 21^+ and ligated to SmaI and EcoRI cut M13mp18. Sequence obtained from both these clones using the usual universal sequencing primer also contributed towards confirming the sequence of the ORF.

The sequence of the ORF encoded by fragment 21^+ differed from that of the kinase deficient strain λNM1210 by the addition of a sequence of six bases, GTCCTG (Figure 3.13). The sequence of the ORF given in Figure 3.14 includes these six bases (746-751, underlined). Appendix I gives the complete sequence of the contiguous EcoRI fragments 21 and 46 with translations of all ORFs and a map of restriction sites. The polypeptide encoded by fragment 21^+ therefore includes an extra Pro-Gly dipeptide as residues 11 and 12 (Figure 3.14). The molecular weight of this polypeptide was calculated from translation of the DNA sequence as 34.6 kd. The mobility of the 33 kd polypeptide encoded by λNM1108 is slightly higher than that of T4 encoded pnk polypeptide in SDS polyacrylamide gels (Figure 3.2). This can probably be accounted

```
      1570                1590                1610                1630                1650                1670
GTAGTTGAAATGTGGCGTCGTATCGGTGTTGAATGCTGGCAAGTCGCTTCGGGAGATTTTTAATGGCTTGGCACCATGAAACTTGGGCTATTGTTATTGTAAATAGCGGTTTAGTTGGTA
ValValGluMetTrpArgArgIleGlyValGluCysTrpGlnValAlaSerGlyAspPheEnd

      1690                1710                1730                1750                1770                1790
CTAGTAATGGGCAATTTTGTGTATTTACTAGTGAAAATAGAGCATGGGAGGAATGTCTTAAATTAAGAGAAAAGAATCCCGATGTTGAACTAGTAGTAAAGAAAACTAAACTGCCTTTAC

      1810                1830                1850                1870                1890                1910
CATGGAAAACTTATGAATAACCTAGAAAAGATTTATCGTCTTTGTGATAAAATTGAAAAAGAAAAGAAATATCTATTTTGTCTATGGCCTATTGTTGACGGAAGAGTAGGCCTAGATGTT

      1930                1950                1970                1990                2010                2030
CTTGATTATGAAACAGAAGACAGAGTAGATGGTTCAACTTTTGATAATGCGTTGGATGTTATTGATTGGCTTGAAGAAAATTATGTGAGGTAAATATGTTTCCGACTTACTCTAAAATCG

      2050                2070                2090                2110                2130                2150
TAGAAGTAGTGTTTAGCCAAATTATCGCTAATAATATGTTTGAAAAACTTGATAACGCAGCTGAGCTTCGAATCCATGCTCAAGTGACTCATGTATTGAACACTTTGCTTCCAGACCAGG

      2170                2190                2210                2230                2250                2270    EcoR1
TGGATTCTATTGCCATTACACTGTATCCAGGTTCCGCGCATATCATTGTCGTATTTGGTCTTGATGCTGAGCTTGTTATCAAAGGCGATATTCGCTTTGAATCTCAAACTGCGGAATTCA

      2290                2310                2330                2350                2370                EcoR1
AAGCGATTTAATGGTTTACTTTACGGTAGAGCTATAATATCACAACTCTACCAAAACAAATGAGGAAAACAAAATGTTGCTAAGTGAAAAACCGATTACTGTTAAAGGAATTC
```

Figure 3.14

Figure 3.14

Sequence of the contiguous EcoRI fragments 21 and 46 with the amino acid sequence of the 906 base ORF encoding pnk polypeptide. The fragment 21 originally isolated in λNM1210 by Mileham et al. (1980) that encoded a 33 kd polypeptide with no kinase activity, did not include the 6 bases GTCCTG (bases 746-751) underlined. The restriction sites for EcoRI, SmaI and XbaI are shown and some possible -10 and -35 promoter sequences are indicated. The -10 region immediately upstream of the ORF appears to lack a -35 region corresponding to an early or middle promoter consensus (Rosenberg and Court, 1979; Brody et al. 1983). The initiation codon is not preceded by a perfect Shine and Dalgarno sequence (Shine and Dalgarno, 1974) but like many ORFs of phage λ (Sanger et al., 1982) by a pronounced purine-rich region. Base 450 remains ambiguous and is read as T or C in independent clones.

for by the difference in their amino acid sequence.

3.1.h. Purification of pnk expressed by λ recombinants and characterisation of its enzymic activities

Active pnk was purified from cells infected with λCM8 and the inactive 33 kd polypeptide was purified from cells infected with λNM1108 by the same procedure (see Section 2.2.v.).

The active protein purified from λCM8 infections was incubated in the presence of [γ-^{32}P]ATP with oligodeoxyribonucleotides generated by digestion of salmon sperm DNA with DNAse I and calf intestinal phosphatase, or with oligoribonucleotides derived from an RNAse T_1 digest of yeast tRNA. Incorporation of ^{32}P into both species (Figure 3.15A) demonstrated that the protein had the expected kinase activity. Subsequent venom digests of the kinase labelled oligonucleotides demonstrated that the label was incorporated into all four deoxyribonucleotides or ribonucleotides (Figure 3.15A).

The 3'-phosphatase activity was detected by the ability to release ^{32}P from 3'-[^{32}P]CMP but not 5'-[^{32}P]CMP. The specifically labelled substrates were isolated from appropriate digests of λcI857 DNA which had been nick-translated in the presence of [α^{32}-P]dCTP (see section 2.2.y). The protein purified from λCM8 infections showed 3'-phosphatase activity, while the defective protein

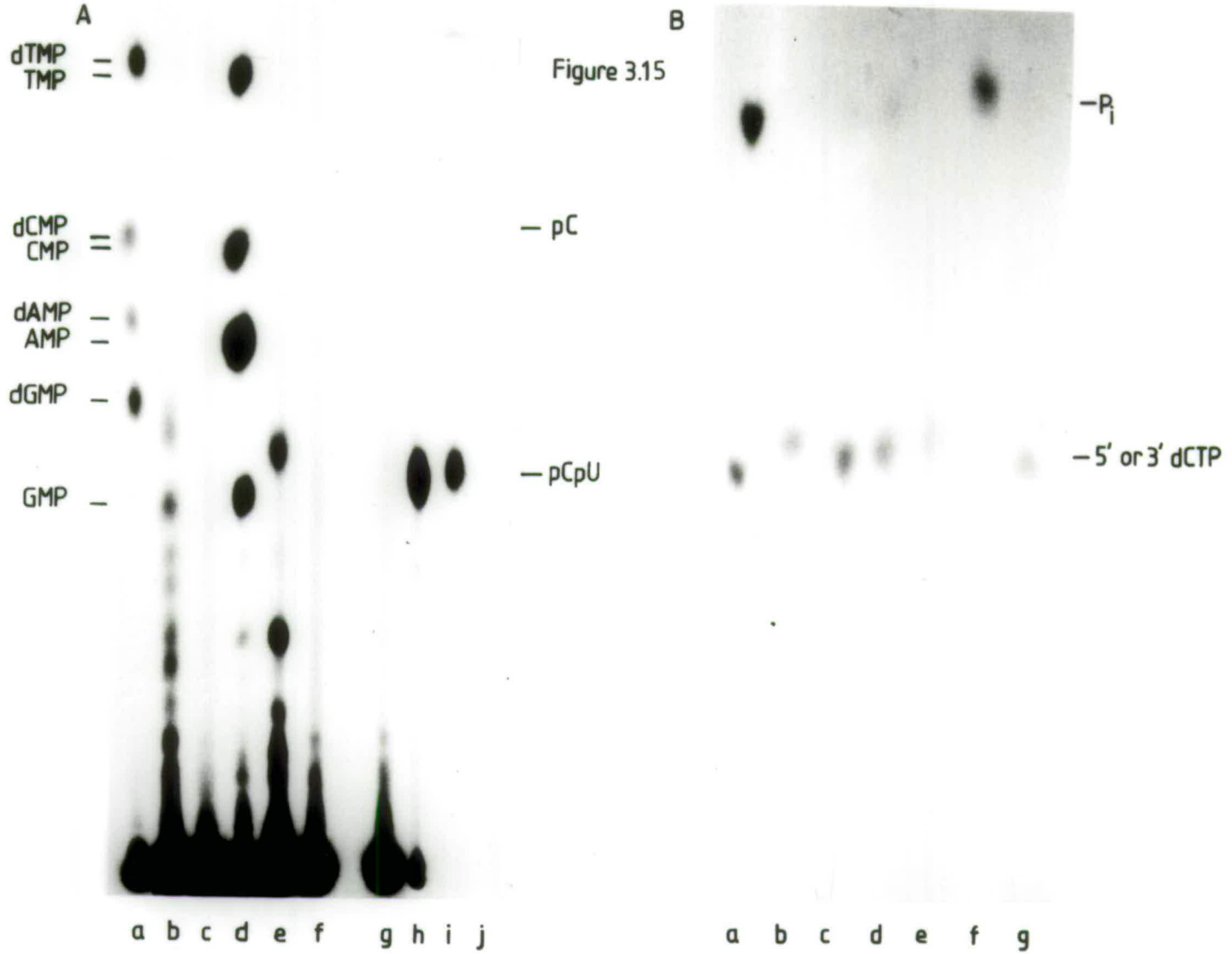

A
dTMP
TMP
dCMP
CMP
dAMP
AMP
dGMP
GMP
— pC
— pCpU
a b c d e f g h i j
B
—Pi
— 5' or 3' dCTP
a b c d e f g

Figure 3.15

Figure 3.15

A. Specificity of the 5'-kinase activity of pnk purified from ED8689 cells infected with λCM8. ^{32}P labelled products of the incubations were separated by ionophoresis on AE81 cellulose paper at pH 3.5. Tracks are: (a) λcI857 DNA digested with pancreatic DNAse (DNAse I) and incubated with pnk and then venom phosphodiesterase. The positions of 5'-dNMPs are indicated; (b) DNA digested with pancreatic DNAse and incubated with pnk; (c) DNA digested with pancreatic DNAse; (d) yeast tRNA digested with RNAse T_1 and incubated with pnk and then venom phosphodiesterase. The positions of 5'-NMPs are indicated; (e) yeast tRNA digested with RNase T_1 and incubated with pnk; (f) yeast tRNA digested with RNase T_1; (g) distilled water incubated with pnk; (h) CpU incubated with pnk; (i) [^{32}P]pCpU eluted from a previous separation by paper ionophoresis; (j) incubation of [^{32}P]pCpU with venom phosphodiesterase. The positions of pCpU and pC are indicated.

B. Specificity of the 3'-phosphatase activity of pnk purified from ED8689 cells infected with λCM8, or of 33kd polypeptide purified from infection with λNM1108. Tracks are: (a) 5'-[^{32}P]dCMP

digested with calf intestinal phosphatase, the position of the released phosphate is indicated; (b) 5'-[^{32}P]dCMP incubated with distilled water; (c) 5'-[^{32}P]dCMP incubated with pnk from the λCM8 infection; (d) 5'-[^{32}P]dCMP incubated with 33 kd protein from the λNM1108 infection; (e) 3'-[^{32}P]dCMP incubated with distilled water; (f) 3'-[^{32}P]dCMP incubated with pnk from the λCM8 infection; (g) 3'-[^{32}P]dCMP incubated with 33 kd protein from the λNM1108 infection.

purified from λNM1108 infections did not (Figure 3.15B). Direct assays on crude lysates were not done since it was assumed that contaminating host phosphatases would obscure the results. Instability of the purified defective protein was therefore not ruled out as a possible cause of 3'-phosphatase inactivity.

A sample of pnk purified from infections with λCM8 was sent to Dr. R. Mattaliano (Biogen, Cambridge, Mass.) for N-terminal amino acid sequence analysis by micro-techniques. These results are presented in the Discussion.

3.2 AMPLIFICATION OF POLYNUCLEOTIDE KINASE PROTEIN

The enzymic activities and substrate specificities of pnk purified from cells infected with λCM8 were indistinguishable from those of the T4 encoded enzyme. However, λCM8 was not a convenient source for routine preparations of the enzyme since high titre phage lysates were required to infect bacteria and infected cells had to be harvested before lysis, i.e. after only 10-15 min of infection. The necessity for phage infections can be avoided in 2 ways; use of an integration-proficient λ phage which is defective in lysis and can be propagated as a lysogen, or use of a high copy number plasmid vector. On the assumption that the pseT gene has been cloned on a fragment including a functional promoter, a high copy number plasmid system may be susceptible to deleterious effects

of constitutive pnk over-expression. It would be necessary to construct a series of fusions of fragment 21^+ in the plasmid vector in order to position the pseT gene downstream of a controllable promoter without the upstream pseT promoter sequences.

The method previously used for T4 DNA ligase (Murray et al., 1979) was chosen as the simpler approach, that is to position fragment 21^+ in the centre of an integration-proficient, temperature-inducible vector that is deficient in both lysis and packaging. Propagation of this phage as a prophage, i.e. as a single copy per host chromosome, and with the major λ promoters repressed should minimise adverse effects of pnk overproduction until the phage is induced. Following induction of such a lysogen the λ genome would replicate to give 100 or more copies per bacterium, and transcription of the amplified pseT gene would result following activation of the late λ promoter $p_{R'}$. This system would be more convenient than infecting with high titres lysates, but would only be successful if the putative promoter upstream of pseT in fragment 21^+ is not sufficiently strong that deleterious levels of pnk are present in the uninduced lysogen.

Fragment 21^+ was purified from an EcoRI digest of λCM8 DNA and ligated to EcoRI cut λNM1070 vector (Wam Eam lacZ att^+ int^+ cI857 Sam). Fragment 21^+ would replace the EcoRI fragment of NM1070 that encodes most of the lacZ polypeptide (Figures 3.10B and 3.12)

The plaques recovered were hybridised to a probe made from purified fragment 21^+ DNA. Positives were checked for λ immunity by plating out on hetero-immune (C600 (λ*imm*^{434})) and homo-immune (C600 (λ)) lysogens, and for inability to grow on a suppressor-free strain (NM594). NM1070 phage including fragment 21^+ always gave minute plaques. Lysates were made, and although their titres were not very high, it was possible to infect ED8689 cells with 5 different lysates and demonstrate the presence of kinase activity in crude cell extracts (Figure 3.16). Infection of UV irradiated cells (M159) with one such phage λCM21 showed expression of the 33 kd pnk polypeptide late in infection (Figure 3.6).

Due to the difficulty of preparing high titre lysates, DNA could not be prepared from infection with e.g. λCM21. Therefore the *supE* host C600 was lysogenised with λCM21 and with a second recombinant, and induction of these lysogens (see sections 2.2.f and 2.2.g) provided phage for DNA extraction. *Sma*I digests of these two recombinant DNAs resulted in loss of the 32.31 kb λNM1070 vector left arm and appearance of 20.73 and 6.98 kb fragments (λCM21 in Figure 3.11) which when compared with *Sma*I digests of λNM1104 and λCM1 in Figure 3.5 show that fragment 21^+ is in the same orientation as fragment 21 in λNM1104. Fragment 21^+ in λCM21 (Figure 3.12) therefore cannot be transcribed from the early λ promoter p_L.

λCM21 includes amber mutations in the *W* *E* and *S* genes and is therefore defective in lysis and packaging.

Figure 3.16

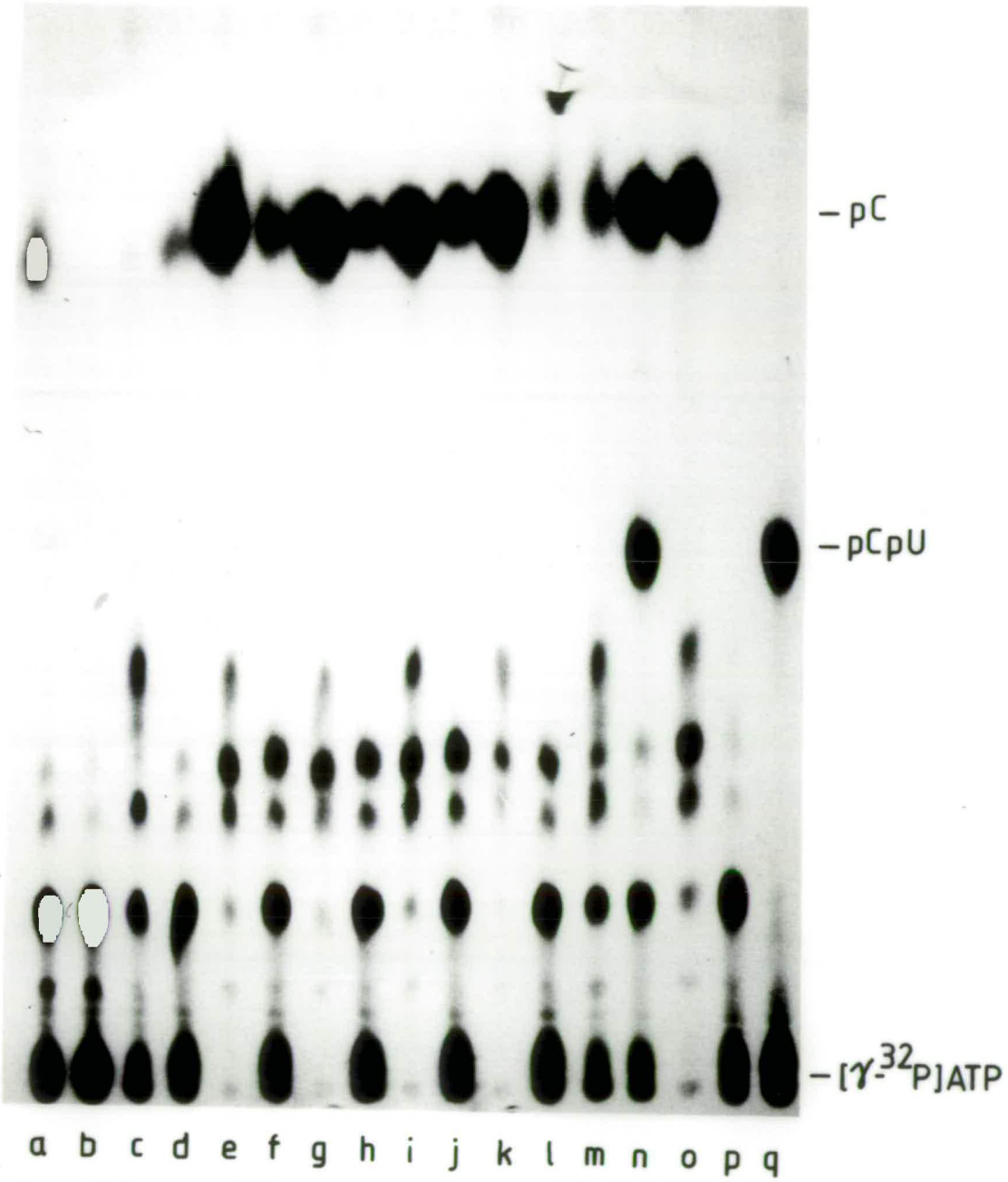


Figure 3.16

Assays for kinase activity in crude cell extracts of ED8689 infected with λ recombinants encoding an active pnk. Radiolabelled products of the kinase reaction were separated by ionophoresis on AE81 cellulose paper at pH 3.5. Tracks are: (a) T4D (wild type); (b) T4pseT2; (c) uninfected cells (ED8689); (d) and (e), (f) and (g), (h) and (i), (j) and (k), (l) and (m), are 5 recombinant derivatives of λNM1070 including fragment 21^+ derived from λCM8. Each was assayed after 15 min of infection (d), (f), (h), (j) and (l) and after 2 h of infection (e), (g), (i), (k) and (m). Track (n) is λCM8; track (o) is a second recombinant derivative of λNM1149 (see Figure 3.10B, track (f)) including fragment 21^+; track (p) is distilled water; track (q) is purified T4 encoded pnk. The positions of [γ-^{32}P]ATP and of the labelled products of the kinase reaction, pCpU or pC are indicated.

In order to employ these features to further amplify pnk expression it was necessary to propagate the phage in a suppressor-free background; the suppressor-free host ED8689 was therefore lysogenised with λCM21. A second lysogen including the vector λNM1070 as the prophage was also isolated. The lysogens were propagated at 32°C and induced by inactivation of the heat-labile cI857 repressor (see section 2.2.u), to achieve the expression of phage encoded polypeptides.

Crude cell extracts derived from induction of the lysogen ED8689 (λCM21) and from infection of ED8689 cells by T4D (wild-type) phage were assayed for kinase activity. The crude lysates were first adjusted to equivalent total protein concentrations as determined by Lowry-Folin assays (see section 2.2.w) and were then further diluted serially. Aliquots of these dilutions were assayed for kinase activity in terms of the conversion of the dinucleoside monophosphate CpU to the dinucleotide pCpU by the transfer of ^{32}P from [γ-^{32}P]ATP. As previously mentioned (section 3.1.a) a nuclease activity in crude cell extracts degrades pCpU to pC. Consequently in the presence of 5'-kinase activity, a radiolabelled spot migrating with the mobility of 5'-dCMP was seen on separation of reaction products by ionophoresis. This pC spot was gradually lost as the T4D crude extract was diluted (Figure 3.17). In contrast, when reactions were carried out with serial dilutions of the induced lysogen ED8689 (λCM21), the pC spot was

Figure 3.17

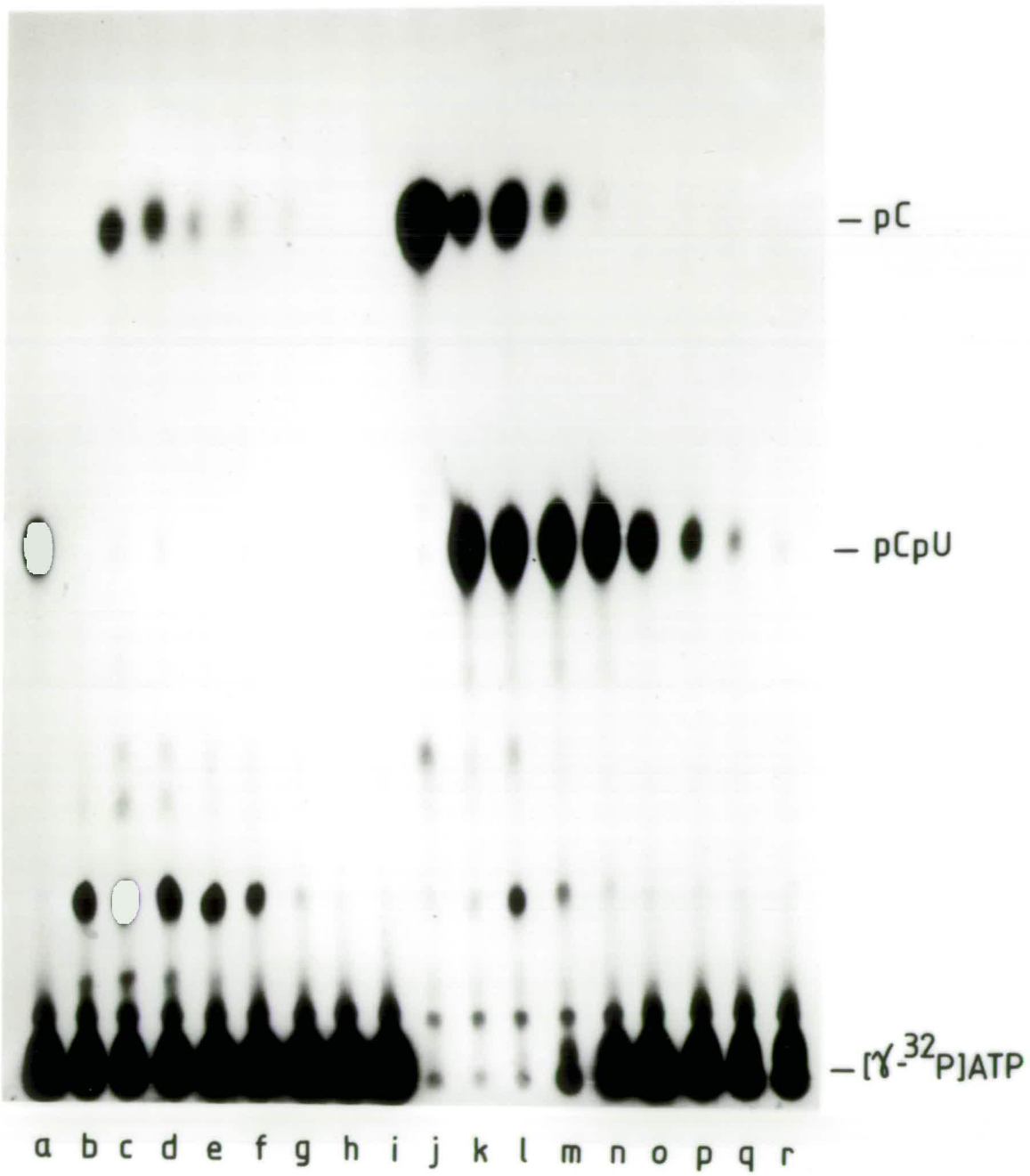


Figure 3.17

Assay for kinase activity in crude cell extracts after infection of ED8689 cells with T4D (wild-type), or after induction of a lysogen ED8689 (λCM21). Radiolabelled products of the kinase reaction were separated by ionophoresis on AE81 cellulose paper at pH 3.5. Tracks are: (a) purified T4 encoded pnk; (b) T4pseT2 (a pseT$^-$ strain); (c) to (i), 2-fold serial dilutions of a crude extract of an infection with T4D; (j) to (r), 3 fold serial dilutions of a crude extract of the induced lysogen ED8689 (λCM21). The undiluted crude extracts in tracks (c) and (j) were adjusted to the same total protein concentration before further dilution. The positions of [γ-^{32}P]ATP and of the labelled products of the kinase reaction, pCpU and pC are indicated.

replaced at increasing dilutions by a spot migrating in the position of pCpU. The pCpU spot was only lost after at least a 2000-fold dilution of the extract. Amplification of pnk activity in the induced lysogen crude extract appeared to be sufficiently great that the pCpU product of kinase activity remained after the contaminating nuclease activity had been lost by dilution. In crude extracts of the induced lysogen (particularly tracks (j)-(l) in Figure 3.17) there was sufficient kinase activity to use up all the available [γ-^{32}P]ATP phosphate donor. In the most concentrated extract (track (j), Figure 3.17) the accumulation of pCpU is presumably prevented in the presence of a high concentration of contaminating nuclease because the availability of phosphate donor limits the 5'-kinase activity. Since the lysogen extract was diluted at least 100 fold more than the T4D extract before the products of kinase activity could no longer be detected, the amplification of kinase activity was 100-fold that in T4D infections.

Expression of pnk polypeptide on induction of the lysogen ED8689 (λCM21) was observed after Coomassie blue staining of polypeptides separated on 15% linear SDS polyacrylamide gels (Figure 3.18A). The lysogen including λNM1070 expressed the *lacZ* polypeptide (β-galactosidase) under the same conditions. In Figure 3.18A, the pnk polypeptide had separated into a doublet. This was rectified in later gels (Figure 3.18B) by increasing the reducing agent (β-mercaptoethanol) 2 fold

Figure 3.18

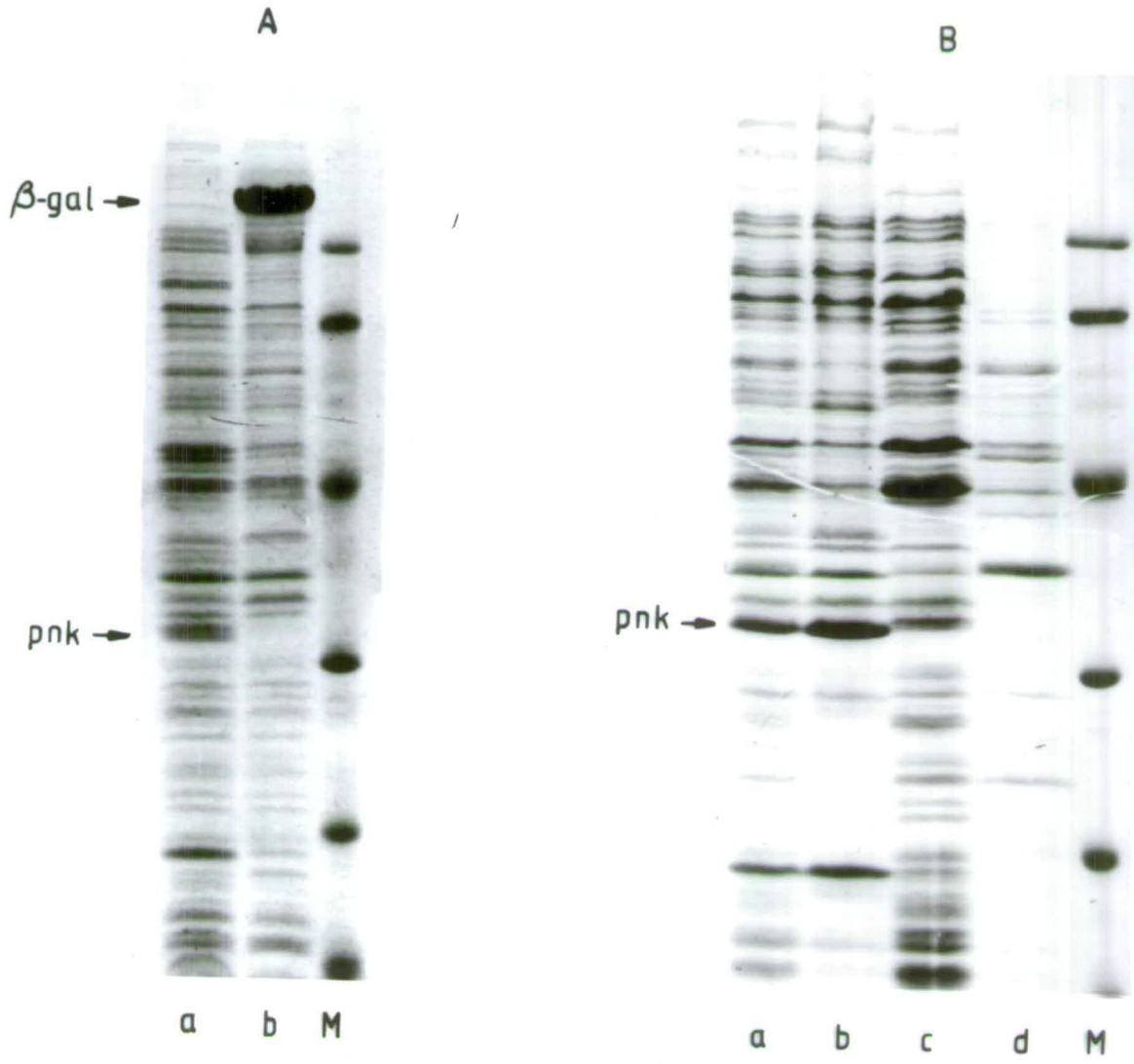


Figure 3.18

A. Analysis of polypeptides after induction of a λ prophage including fragment 21^{+}. The polypeptides were separated on a 15% linear SDS polyacrylamide gel and stained with Coomassie blue. Tracks are: (a) crude extracts from an induced lysogen ED8689 (λCM21); (b) crude extracts from an induced lysogen ED8689 (λNM1070); (M) standard marker proteins (see legend to Figure 3.2) giving marker sizes (kd) of 94, 67, 43, 30, 20.1 and 14.4 (α-lactalbumen). The positions of β-galactosidase expressed by vector λNM1070 and of pnk expressed by λCM21 are indicated. In this gel pnk migrates as a broad (doublet) band due to insufficient reducing agent.

B. Analysis of polypeptides at stages during the purification of pnk from crude cell extracts (section 2.2.v) after induction of a lysogen ED8689 (λCM21). The polypeptides were separated on a 15% SDS polyacrylamide gel and stained with Coomassie blue. Tracks are: (a) total protein in a crude cell extract after induction of the lysogen; (b) protein precipitated from the crude cell extract in 35% ammonium sulphate; (c) protein precipitated from the crude cell extract in 35-55% ammonium sulphate; (d) protein remaining in the supernatant after 35-55% ammonium sulphate precipitation; (M) standard marker proteins (see above). The position of pnk is indicated.

in the loading dye. Figure 3.18B also shows the location of pnk polypeptide in the initial stages of its purification from the induced lysogen ED8689 (λCM21) (see section 2.2.v). The majority of pnk polypeptide was precipitated in 35% ammonium sulphate. Accurate measurements of protein concentration in purified protein preparations were not made. However, some idea of the pnk level as a percentage of total protein in the induced lysogen was gained from a densitometer scan of separated polypeptides in a 15% linear SDS polyacrylamide gel (Figure 3.18B). After Coomassie blue staining, individual tracks in the gel were subjected to scanning with a Fisons Vitatron TDL-100 densitometer. A scan of separated total protein (Figure 3.18B track (a)) is shown in Figure 3.19 where peaks represent quantitatively the polypeptide bands. Comparison of the area under the peak corresponding to the pnk polypeptide band with the total area under the curve, indicated that pnk polypeptide constituted at least 7% of the total soluble cell protein in this crude extract of the induced lysogen ED8689 (λCM21).

In order to investigate the stability of the lysogen ED8689 (λCM21) cells were subcultured through seven successive cycles. Each time, an overnight was diluted 1000 fold, grown to saturation at 32°C and used to inoculate a new overnight. Aliquots of the culture at each stage were plated at 32°C and 42°C. After seven cycles, over 90% of bacteria remained lysogenic, i.e. did not give colonies at 42°C. Nine of these lysogenic colonies were

Figure 3.19

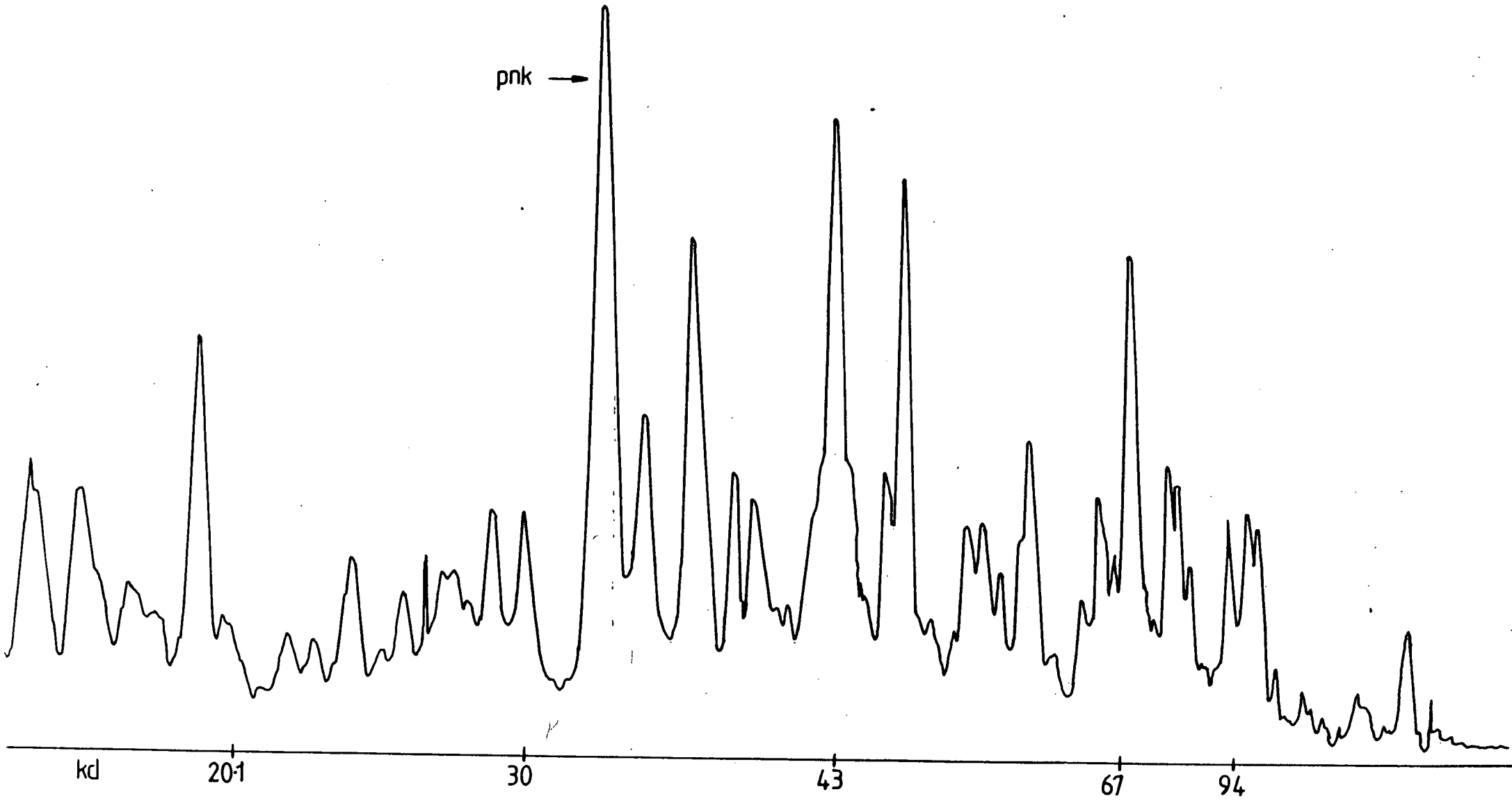


Figure 3.19

A trace representing total protein in a crude lysate after induction of ED8689 (λCM21). This was derived from a densitometer scan of polypeptide bands in Figure 3.18B, track (a). The positions of standard marker proteins are shown. Comparison of the total area under the curve and of the area of the peak corresponding to the pnk polypeptide band indicates that pnk constitutes about 7% of total cell protein in this crude extract.

grown up and induced and crude cell lysates were tested for kinase activity. All gave high levels of activity. This indicates that the prophage is relatively stable and it should be possible to maintain ED8689 (λCM21) as a lysogen for long periods.

C H A P T E R 4

DISCUSSION

4.1 CLONING AND EXPRESSION OF PNK

Since the pseT gene is non-essential it was necessary to identify clones by their ability to confer kinase activity on host cell extracts. The identification of the coding sequence for pnk was impeded by the fact that T4 EcoRI fragment 21 cloned in a λ vector by Mileham et al. (1980) carried a defective pseT gene. Recombinants including fragment 21 expressed a polypeptide of 33 kd, the appropriate size for the pnk subunit, but did not induce kinase activity. In infections with T4 phage, only one 33 kd polypeptide could be detected that was missing in infections with pseT2 ($pseT^-$) phage. This suggested that the 33 kd polypeptide encoded by T4 EcoRI fragment 21 was the pnk subunit. The 33 kd polypeptide encoded by fragment 21 seemed to have a slightly higher mobility than T4 encoded pnk polypeptide on SDS polyacrylamide gels. This raised the possibility that the sequence encoding the pnk subunit may extend outside fragment 21, although a lack of essential processing or a mutation in the coding sequence could be the cause of inactivity.

DNA sequencing of fragment 21 revealed an ORF of 900 bases in the orientation required for expression of the 33 kd polypeptide. The sequence did not extend into the flanking fragments. The position of the SmaI site 75 base pairs downstream of the start of the ORF was consistent with the lack of expression of the 33 kd

polypeptide in cells infected with λCM6. These data suggested that the ORF within fragment 21 encoded the 33 kd polypeptide identified in infections with λ recombinants. No obvious reason for the inactivity of the polypeptide could be deduced from the DNA sequence, except that the sequence was not consistent with a report that the N-terminal residue of pnk was phenylalanine (Lillehaug, 1977). A phenylalanine occurred at position 20 of the 33 kd polypeptide, but a difference in size of 19 amino acid residues (about 2.2 kd) between the T4 encoded pnk polypeptide and the 33 kd polypeptide should have been noticeable on SDS polyacrylamide gels. Processing by the addition of an N-terminal phenylalanine may occur.

The pseT gene has been mapped downstream of the alc/unf gene (Sirotkin et al., 1978) and since the T4pseTΔ1 strain was reported to be alc$^-$, it was assumed that there was unlikely to be an ORF large enough to encode the pnk subunit between alc/unf and fragment 21. Unfortunately, repeated attempts to obtain recombinants including DNA downstream of fragment 21 were unsuccessful and the possible existence in this region of other ORFs capable of encoding a 33 kd polypeptide was not ruled out.

The possibility of a small point mutation or deletion within the coding sequence lead to the recloning of fragment 21$^+$ derived from a different source of T4 DNA. It was thought that mutations which would alleviate

deleterious effects of pnk expression or activity may be preferentially selected in cloned sequences. However, a new clone including fragment 21^+ gave considerable kinase activity in T4 infections, and the DNA sequence of this fragment revealed the presence of an extra 6 base pair sequence within the ORF. Fragment 21^+ purified from this recombinant, λCM8, was transferred to the central region of a λ vector (λNM1070) where expression would be enhanced by transcription from λ promoters. All such recombinants gave very small or minute plaques, many of which may have escaped detection. Expression of active pnk from λ promoters did seem to be deleterious to phage propagation. Consistent with this was an apparent bias against cloning fragment 21 in the orientation in which pnk expression would be enhanced by transcription from the early λ promoter p_L. Even for those phage which were recovered, for example, λCM21, where expression of pseT was enhanced by transcription from $p_{R'}$, plaques were minute and lysates were difficult to prepare by infection. A high titre of λCM21 phage for DNA preparations was only achieved following induction of a supE lysogen including a λCM21 prophage and concentration of the cells before lysis.

It would appear from the unexpectedly high expression of pseT in λNM1149 recombinants, and the high background of early expression where transcription should be mediated exclusively from the λ late promoters, that fragment 21 includes a promoter upstream of pseT.

However, experiments *in vitro* have not detected a strong promoter in this region (Gram *et al.*, 1984). The DNA sequence in Figure 3.14 shows some possible early promoter sequences upstream of the *pseT* coding sequence. A -10 region occurs immediately upstream of the ATG, however, there is no obvious sequence relating to either an early or a middle type -35 promoter consensus sequence. One alternative possibility is that *pseT* is co-transcribed from a promoter sequence situated upstream of an ORF of 228 bases which stops adjacent to the *pseT* initiation codon (Figure 3.9 and Appendix I). The sequence immediately upstream of *pseT* does not possess a perfect Shine-Dalgarno consensus sequence, GAGGA (Shine and Dalgarno, 1974), however, *pseT* like many ORFs of phage λ (Sanger *et al.*, 1982) is preceded by a purine-rich region.

Propagation of λCM21 as a prophage, i.e. as a single copy per host chromosome, and with the major λ promoters repressed, should provide a means of stabilising the coding sequence i.e. of suppressing deleterious levels of pnk expression until expression is induced allowing excision and multiplication of the phage and enhancement of *pseT* transcription from λ promoters. The suspected presence of a promoter on fragment 21 was one reason for not using a plasmid expression system. Unless the promoter was removed before introducing the fragment into a high copy number vector, constitutive expression of pnk may have prevented recovery of clones including an intact *pseT* sequence.

There is no apparent explanation for the occurrence of a 6 base pair deletion, GTCCTG, in the original fragment 21 clone (λNM1210). The surrounding DNA sequence seems unlikely to include secondary structures capable of affecting replication or recombination. Although the deletion results in loss of 5'-kinase and apparently also 3'-phosphatase activity, it does not relieve the deleterious phenotype associated with fragment 21. It was, for example, never possible to lysogenise *E.coli* with an integration-proficient phage carrying the defective *pseT* gene (e.g. λNM1104), and it was difficult to grow and maintain a culture of *E.coli* carrying a plasmid including the defective gene (N. Murray, personal communication).

4.2 THE STRUCTURE OF THE PNK POLYPEPTIDE

The amino acid sequence encoded by the cloned *pseT*$^+$ gene was searched for homology with protein sequences in the NBRF protein sequence database (National Biomedical Research Foundation, Georgetown University Medical Centre, Washington, D.C.). This revealed a striking homology between the N-terminal region of the *pseT* sequence and the N-terminal region of pig and human adenylate kinase (Table 4.1A), an enzyme that catalyses the phosphorylation of AMP by ATP. Crystallographic analysis of various kinases has shown a common structural feature; the enzymes contain two lobes separated by a cleft (Anderson *et al.*, 1979). The substrates bind in the cleft, and in hexokinase and adenylate kinase (Pai *et al.*, 1977) the cleft is known to

Table 4.1

	Protein	Residues	Sequences
A.	E.coli ATPase β	143-165	G G (K) V G L F [G] G A G V [G K T] V N M M E L I R
	E.coli ATPase α	162-184	G Q (R) E L I I [G] D R G [T G K T] A L A I D A I I
	RecA-protein	58- 84	M G (R) I V E I Y [G] P E S [S G K T] T L T L Q V I A
	Adenylate kinase	7- 30	K S (K) I I F V V [G] G P G [S G K] G T Q C E K I V Q
	Active pnk	1- 24	M K (K) I I L T I [G] C P G [S G K S] T W A R E F I A
	Inactive pnk	1- 22	M K (K) I I L T I [G] C [S G K S] T W A R E F I A
B.	E.coli ATPase β	227-249	M A E K F [R] D E [G] R D V [L L F V] (D) N I Y R Y T
	E.coli ATPase α	265-287	M G E Y F [R] D R [G] G D A [L I I Y] (D) D L S K Q A
	Adenylate kinase	102-124	G E E F E [R] K I [G] Q P T [L L L Y] V (D) A G P E T
	Phosphofructokinase	85-107	G I E Q L [K] K H [G] I Q G [L V V I] G G (D) G S Y Q
C.	Histone 2b	100-123	L P G E L A K H A V S E [G T K A V] T K Y T T S K
	Phosphoglycerate kinase	338-363	P V G V F E W E A F A R [G T K A L] M E D V V K A T S
	Inorganic phosphatase	96-121	P Q T W E D P N V S H P E [T K A V] G D N N P I D V L
	Active pnk	115-140	P W T E L V K R N S K R [G T K A V] P I D V L R S M Y

Table 4.1

A. shows alignment of homologous sequences in active and inactive pnk polypeptides and other adenine nucleotide binding proteins. The conserved feature; GXXXXGK(T)XXXXXXI/V, usually preceded by a basic amino acid (encircled) is indicated by boxes.

B. shows the second region of conserved sequence associated with adenine nucleotide binding in some proteins. The boxes contain conserved basic or conserved hydrophobic residues. The conserved aspartic acid residue is encircled. This conserved region is not present in pnk.

C. shows a region of pnk which shares homology with certain proteins. The significance of this conserved region is not clear. Table 4.1 parts A and B are modified from Walker et al. (1982).

narrow or close upon binding the substrates. This change in conformation seems to be important for enzyme specificity and for the mechanism of kinase activity. Analogues of glucose with bulky side chains are competitive inhibitors of the binding of glucose to hexokinase, but they are not substrates for the enzyme since although they bind in the correct orientation the bulky side chains prevent closing up of the two lobes of the enzyme (Anderson et al., 1979).

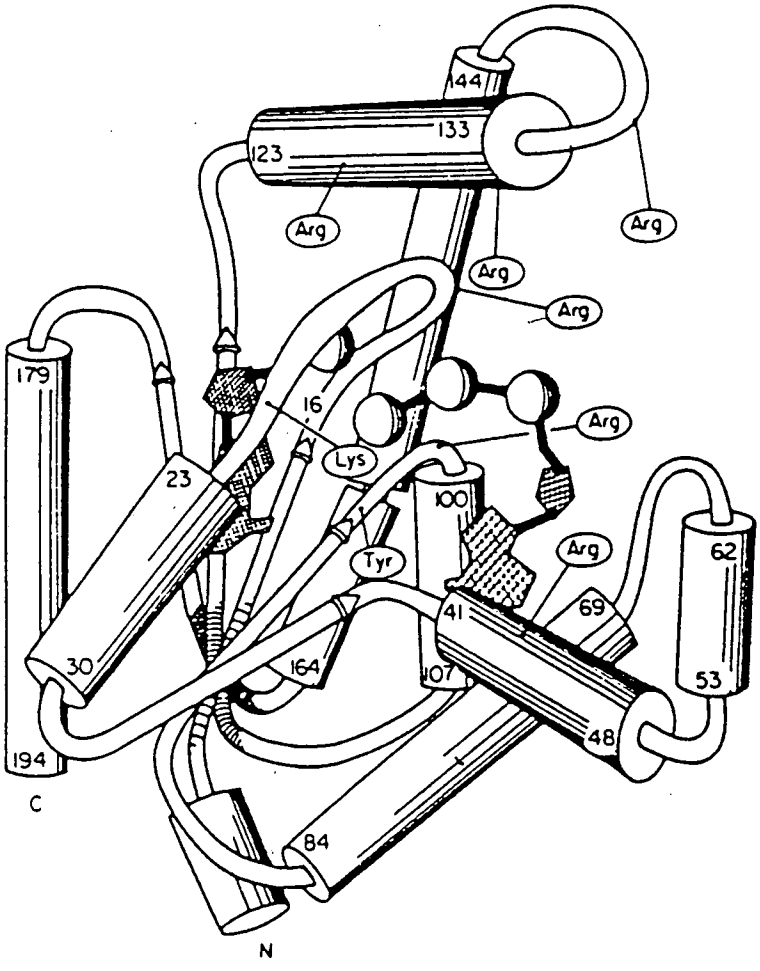
Adenylate kinase is a 22 kd monomer and again the active site is contained within a cleft formed between the two lobes of the enzyme. AMP and ATP bind at opposite ends of the cleft with their phosphates extending towards each other in the centre of the cleft (Pai et al., 1977). Two markedly different confirmations of crystalline adenylate kinase have been described (Sachsenheimer and Schulz, 1977). In the A conformation the cleft has been narrowed by a few angstroms compared to the B conformation. This motion seems largest at the proposed AMP binding site, and Pai et al. (1977) concluded that the open conformation B is related to the native conformation of the enzyme, and that AMP binding causes the protein to switch to the A conformation in which the cleft is closed up.

Crystallographic analyses of several kinases with diverse secondary and tertiary structures show the same bilobal character, but in most cases the enzymes have not been crystallised in the presence of the substrates. However, the substrate-induced closing of the cleft between

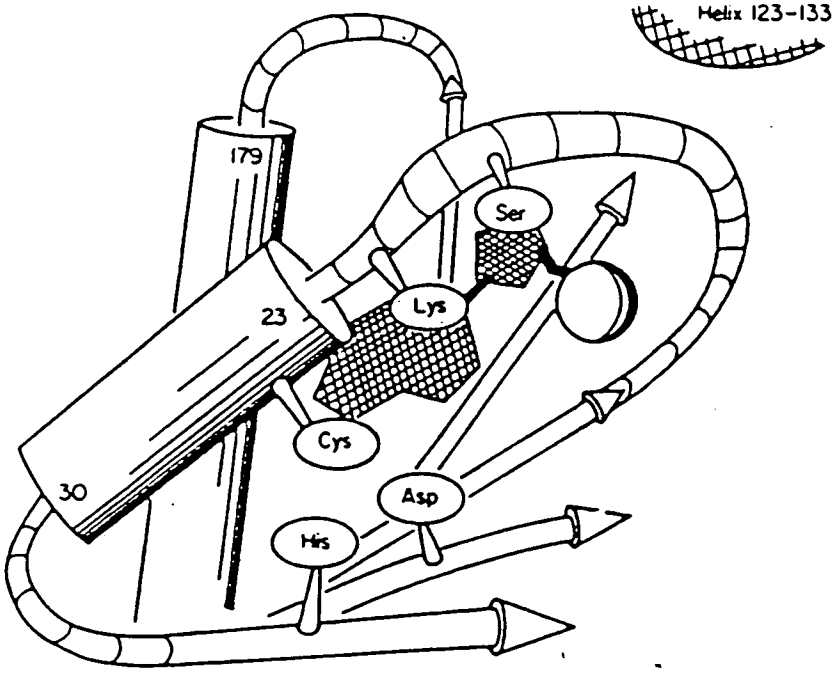
the two lobes is probably a common feature of kinase enzymes. This conformational change may be required to 'embrace' the substrate and orient catalytic groups, or to exclude solvents and promote nucleophilic attack by the phosphate group acceptor on the phosphate group to be transferred.

The substrates of adenylate kinase are the adenine nucleotides, AMP and ATP. According to kinetic experiments these substrates occupy two distinct positions, the AMP site and the ATP site (Noda, 1958; Rhoads and Lowenstein, 1968). The AMP site is very specific and the adenine is buried deep inside the protein, whereas the ATP site is less specific and can bind other nucleotides (Secrist et al., 1972; Slotin and Hampton, 1975). The most essential difference between the A and B forms of crystalline adenylate kinase (Sachsenheimer and Schulz, 1977) is the opening of a pocket behind loop 16-22 and helix 23-30 (Figure 4.1A) which is large enough to accommodate adenosine. The His36 at the active site of the enzyme (Cohn et al., 1971) is located at the entrance of this pocket. Pai et al. 1977 have shown that the B and A forms bind AMP and ATP respectively and are related to the inactive E conformation of the enzyme before and the active E* conformation after the induced-fit transition postulated by Koshland (1958) and Jencks (1975). Hence the B conformation (the native enzyme) has an open ATP site, a wide cleft for accommodating phosphates and

Figure 4.1


144
133
123
Arg
Arg
Arg
Arg
179
16
Lys
Arg
23
100
Tyr
Arg
62
69
41
30
164
107
53
48
194
C
84
N

(a)


Helix 123-133
179
Ser
Lys
23
Cys
30
Asp
His

(b)

Figure 4.1

Sketches of adenylate kinase from Pai et al. (1977). (a) represents the 3-dimensional structure of the whole protein. Helices are represented by cylinders, and the five strands of β-pleated sheet by arrows. The substrates AMP and ATP are shown bound to the enzyme. The side chains of Lys21, Arginines 44, 97, 128, 132, 138, 149 and Tyr95 are shown. (b) shows adenylate kinase in crystal form B with a view into the deep pocket which has been assigned to the AMP binding site. The positions of Ser19, Lys21, Cys25, His36 and Asp93 are indicated. This conformation of the enzyme has been related to the free enzyme. Crystal form A has been related to the active enzyme, E*, which has undergone an induced-fit conformational change. In crystal form A the AMP binding pocket is closed by movements of loop 16-22 and helix 23-30.

an open AMP site (Figure 4.1). The A conformation also has an open ATP site, a narrow phosphate cleft, and the adenosine pocket at the AMP site is closed. This corresponds to an induced-fit model where binding of AMP triggers the switch between these two quite stable conformations (Sachsenheimer and Schulz, 1977).

The specific binding positions of ATP and AMP in pig adenylate kinase have been located by X-ray diffraction analysis which identifies the two adenosine pockets and the region of phosphate binding. The ATP binding site is between helices 69-84 and 100-107 (Figure 4.1A) and is lined by Val67, Leu69, Val72, Leu73 and Leu76 on one side and Ile92, Tyr95, Arg97, Gln101 on the other (Pai et al., 1977). Model building suggests that Tyr95 is hydrogen bonded to the N atom of ATP and the 2'-OH and 3'-OH groups to Gln101 and Val67 respectively. The pocket is open to the right and at the bottom (Figure 4.1A) which may explain its non-specificity.

The phosphates of the substrates bind in conformation A to a site lined by numerous positively charged side chains, namely Lys21 and Arginine residues 44, 97, 128, 132, 138 and 149, which probably interact with the negative charges on the phosphate groups.

The large hydrophobic pocket which opens in the B conformation and is assigned to the binding of the adenosine moiety of AMP, is formed between the β-sheet, loop 16-22, helix 23-30 and the C-terminal helix (Figure

4.1B). This pocket is lined by Leu99, Ile11, Val13, Leu16 and Leu18 on the β-sheet, ser19 and Gly20 on the loop and Gln24, Ile28 on one helix and Val86 on the other. The pocket is only accessible in the B conformation (Figure 4.1B) since in the A conformation the entrance is closed by a movement of loop 16-22 (Sachsenheimer and Schulz, 1977). The amino acid sequence of loop 16-22 is Gly-Pro-Gly-Ser-Gly-Lys-Gly, (von Zabern et al., 1976) i.e. every second position is a glycine residue, which allows a large range of bond angles around the glycine carbon atom. Consequently loop 16-22 is very flexible and can perhaps fold around the incoming substrate.

The region of homology between the amino acid sequences of pnk and adenylate kinase involves loop 16-22. In addition this region conforms to a consensus sequence:

GXXXXGK(T)XXXXXXI/V, usually preceded by a basic amino acid, which is conserved in a number of other adenine nucleotide binding proteins (Walker et al., 1982) e.g. E.coli ATP synthases (Table 4.1A). Interestingly the deletion of a 6 base pair sequence in the ORF of the original cloned fragment 21 which expressed inactive pnk, has resulted in the loss of a Pro-Gly dipeptide within the region homologous to loop 16-22 of adenylate kinase (Table 4.1A). This deletion would seriously limit the flexibility of such a loop formed in pnk. If substrate

binding to pnk is analogous to AMP binding to adenylate kinase, it is possible that either the substrate molecule cannot bind to the defective pnk, or that the induced conformational change involving movement of a flexible loop cannot occur and the active E* enzyme conformation cannot be achieved. A lysine immediately following loop 16-22 in adenylate kinase (Lys 21) is conserved in all examples of the consensus sequence proposed by Walker et al. (1982) (Table 4.1A) and this residue probably interacts with the α-phosphate of AMP (Pai et al., 1977). Loop 16-22 is assumed to encircle this phosphate when AMP is bound.

A second region of adenylate kinase which includes residues 110-120 is involved in AMP specific binding. Again this region conforms to a consensus sequence: R/KXXGXXXL-hydrophobic-hydrophobic-hydrophobic-hydrophobic followed by an aspartic acid residue, which is observed in a number of adenine nucleotide binding proteins including phosphofructokinase and E.coli synthases (Table 4.1B). The aspartic acid residue (Asp119 in adenylate kinase) following the consensus sequence is particularly well conserved and may be involved in binding magnesium in adenylate kinase (Walker et al., 1982). This consensus sequence is not present in pnk. This may reflect the difference in substrate specificity since pnk must bind substrate molecules which adenylate kinase cannot i.e. polynucleotides. The search for amino acid sequence homologies with pnk did however

reveal a new consensus including residues 127-131 of pnk (Table 4.1C) which is shared by human and horse phosphoglycerate kinase and yeast inorganic pyrophosphatase. The same sequence is also conserved in histone 2b from various species. The possible relevance of this conserved region remains to be elucidated. Phosphoglycerate kinase and inorganic pyrophosphatase both have phosphatase-like activities. Phosphoglycerate kinase catalyses the phosphoryl transfer reaction:

$$\text{1,3-diphosphoglycerate} + \text{ADP} \overset{Mg^{++}}{\rightleftharpoons} \text{3-phosphoglycerate} + \text{ATP}$$

(Banks et al., 1979), while inorganic pyrophosphatase catalyses the hydrolysis of pyrophosphate to orthophosphate (Cohen et al., 1978). The conserved peptide sequence may be related to 3'-phosphatase activity in pnk, or again it may be involved in substrate binding. The relevance of the histone 2b sequence is not clear.

A topological comparison of adenylate kinase with several dehydrogenases (Schultz and Schirmer 1974) suggested that substrate binding sites in adenylate kinase are equivalent to the NAD binding site in dehydrogenases, with ATP corresponding to the adenosine moiety of NAD. Tyr95 in the ATP pocket (Figure 4.1A) is equivalent to Tyr85 of lactate dehydrogenase and probably both bind to the adenine. There is also a close correspondence in the loop 16-22 region. In lactate dehydrogenase (Moras et al., 1975) the phosphate moiety of NAD is hydrogen bonded to the backbone of an equivalent loop, and in flavodoxin

(Burnett et al., 1974) the loop wraps around the phosphate moiety of FMN. However, Walker et al. (1982), were unable to find sequence homologies with dehydrogenases thought to show structural homologies with adenylate kinase in the region of the nucleotide binding site. This suggests that some structural homologies are related to efficiency of substrate binding rather than the result of divergent evolution.

This thesis gives no experimental evidence that the lack of 5'-kinase and possibly also 3'-phosphatase activity in the defective pnk is the result of the disruption of a loop necessary for substrate binding, or for induced-fit conformational changes. The overall conformation of the enzyme seems unlikely to be markedly different from that of the active enzyme. Both proteins were successfully purified by the same method (see section 2.2.v), so presumably the subunits of the defective kinase associate into a tetrameric species. A comparison of computer predictions of the secondary structure of the pnk polypeptide and defective 33 kd polypeptide is shown in Figure 4.2 (L. Sawyer, personal communication). Residues 10-16, which may be involved in a loop structure analogous to loop 16-22 in adenylate kinase, occur in a region where turns are predicted in the secondary structure of the active pnk polypeptide (Figure 4.2B). Proline is a so-called helix-breaker amino acid residue since its nitrogen atom is in a rigid ring system and the permissible

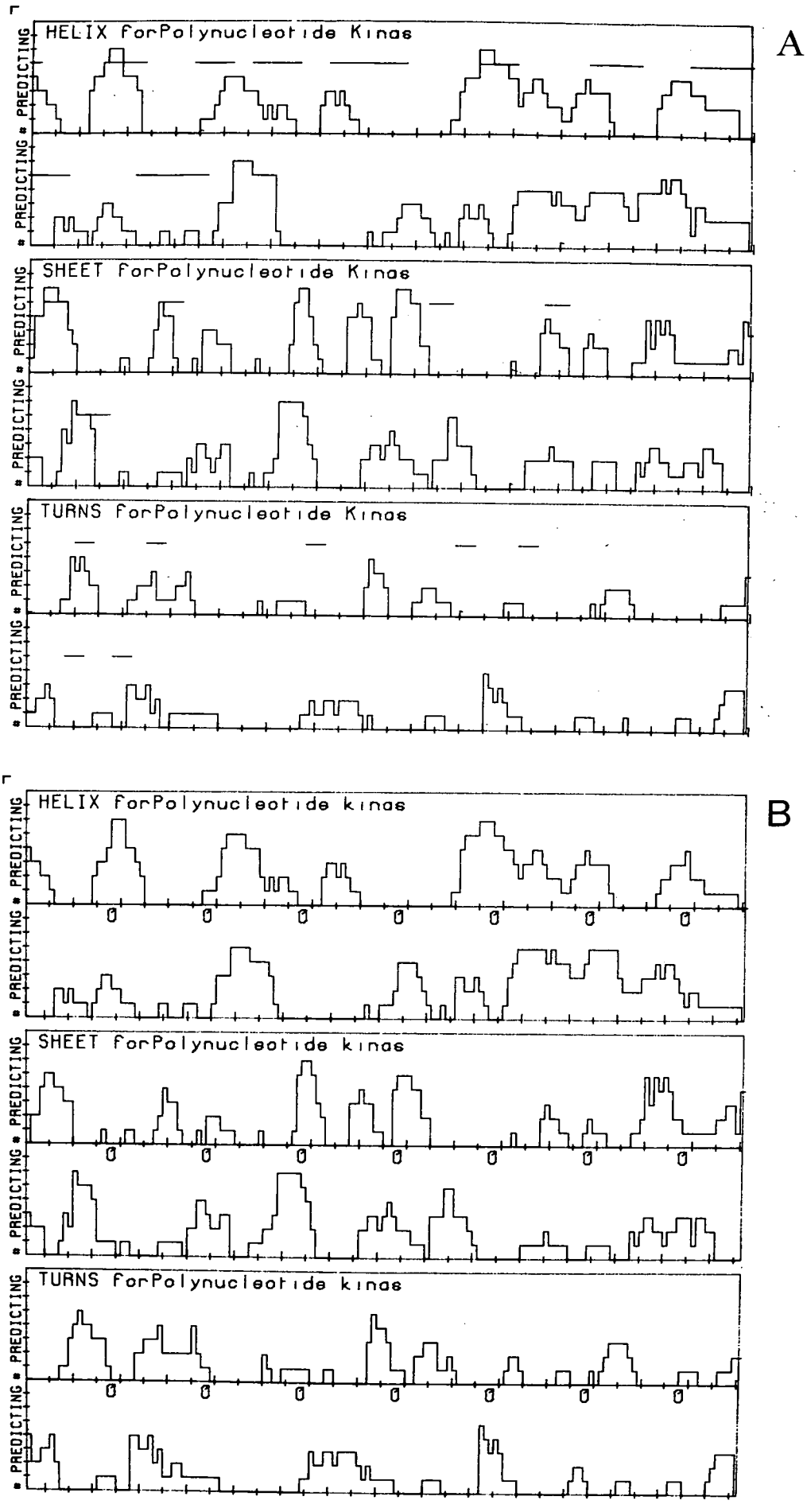

A
HELIX forPolynucleotide Kinas
SHEET forPolynucleotide Kinas
TURNS forPolynucleotide Kinas
B
HELIX forPolynucleotide kinas
SHEET forPolynucleotide kinas
TURNS forPolynucleotide kinas
PREDICTING

Figure 4.2

Figure 4.2

A computer generated histogram showing predictions for regions of α-helix, β-pleated sheet and turns in the peptide sequences of the active and inactive pnk polypeptides encoded by cloned genes (L. Sawyer, personal communication). Marks along the sequence are at intervals of 5 residues starting at the N-terminus.

A. shows the defective pnk polypeptide. Horizontal lines superimposed on the histogram represent the extent of α-helix, β-sheet and turns in the secondary structure of adenylate kinase.

B. shows the active pnk polypeptide. Residues 10-16 thought to be involved in forming a flexible loop occur in a region where turns are predicted in the active pnk polypeptide. This region in the defective polypeptide shows a reduced probability of turns in the polypeptide chain.

rotation around the alpha carbon-nitrogen axis does not allow an α-helical structure, but causes a sharp bend in the polypeptide chain. Hence loss of a Pro-Gly dipeptide (residues 11 and 12 of active pnk polypeptide) in the defective pnk polypeptide reduces the probability of turns in this region (Figure 4.2A).

The DNA sequence of the cloned *pseT*$^{+}$ gene is not consistent with reports that the N-terminal amino acid of pnk is phenylalanine (Lillehaug, 1977). The first phenylalanine encoded by the DNA sequence is at position 22 in the active pnk polypeptide. It is unlikely that 21 N-terminal residues would be lost by processing since the mobilities of the polypeptides of active and inactive pnk are almost identical, and in addition this region includes the putative adenine nucleotide binding site.

A sample of pnk purified as previously described (Section 2.2.v) from infection of a culture of ED8689 with λCM8 was further purified by electroelution of the 33 kd pnk polypeptide band from an SDS polyacrylamide gel. Dr. R. Mattaliano of the Biogen Research Corporation, Cambridge, Mass. USA, carried out peptide sequence analysis of the N-terminal region using this small sample. 20 cycles of analysis gave the following sequence:

N-terminal/Met-Lys-Lys-Ile-Ile-Leu-Thr-Ile-Gly-X-Pro-Gly-Ser-Gly-Lys-Ser-Thr-Trp-Ala-Arg.

No score was obtained for cycle 10 but since cysteine is the only residue not identified in this technique it is inferred in this position by default. Clearly the peptide

sequence of the active pnk expressed by the cloned pseT gene is identical to that predicted by the DNA sequence. The purification of sufficient quantities of pnk from T4 infected cells to allow a similar N-terminal sequence analysis is currently in progress. However, if the T4 protein is modified by addition of an N-terminal phenylalanine, this event is not necessary for 5'-kinase or 3'-phosphatase activity.

The level of enzyme activity produced by the λCM21 prophage system is more than 100 times that in extracts of $T4^+$ infected cells, and for the first time pnk can be prepared without making high titre phage lysates. The four T4 encoded enzymes commonly used in the laboratory: DNA ligase (Murray et al., 1979), RNA ligase (Rand and Gait, 1984), DNA polymerase (W. Konigsberg, personal communication) and polynucleotide kinase can now be prepared without recourse to T4 infected cells. The expression of pnk in the absence of contaminating T4 nucleases is a great aid to purification, and use of currently available gel filtration procedures should allow purification of large quantities of homogeneous protein. A more specific analysis of pnk in vitro in terms of reaction kinetics, substrate binding and structural analysis by crystallography may help in the elucidation of the relationship between the 5'-kinase and 3'-phosphatase activities of the enzyme.

REFERENCES

Adhya, S., Gottesman, M. and de Crombrugghe, B. (1974) Proc.Natl.Acad.Sci.USA, 71, 2534-2538.

Alberts, B. M., Morris, C. F., Mace, D., Sinha, N., Bittner, M. and Moran, L. (1975). In Goulian, M. and Fox, C. W. (eds.) DNA Synthesis and its Regulation, W. A. Benjamin Inc. pp 241-269.

Altman, S. and Lerman, L. S. (1970) J. Mol.Biol., 50, 235-261.

Anderson, C. M., Zucker, F. H. and Steitz, T. A. (1979) Science, 204, 375-380.

Appleyard, R. K. (1954) Genetics, 39, 440-459.

Appleyard, R. K., McGregor, J. F. and Baird, K. M. (1956) Virology, 2, 565-574.

Arber, W. (1983) In Hendrix, R. W., Roberts, J. W., Stahl, F. W., Weisberg, R. A. (eds.) Lambda II, Cold Spring Harbor Laboratory, pp. 381-394.

Banks, R. D., Blake, C. C. R., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., Merret, M. and Phillips, A. W. (1979) Nature, 279, 773-777.

Bayer, M. E. (1968) J.Virol., 2, 346-356.

Becker, A. and Gold, M. (1975) Proc.Natl.Acad.Sci.USA, 72, 581-585.

Becker, A. and Hurwitz, J. (1967) J.Biol.Chem., 242, 936-950.

Belfort, M., Pederson-Lane, J., West, D., Ehrenman, K., Maley, G., Chu, F. and Maley, F. (1985) Cell, 41 375-382.

Bernard, H. U., Remaut, E., Hershfield, M. V., Das, H. K., Helinski, D. R. Yanofsky, C. and Franklin, N. (1979) Gene, 5, 59-76.

Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) Proc.Natl.Acad.Sci.USA, 80, 3963-3965.

Boezi, J. A., Armstrong, R. L. De Backer, M. (1967) Biochem. Biophys. Res. Commun. 29, 281-287.

Brody, E. N., Black, L. W. and Gold, L. M. (1971) J.Mol.Biol., 60, 389-393.

Brody, E., Rabussay, D. and Hall, D. H. (1983) In Matheas, C. K., Kutter, E. M., Mosig, G. and Berget, P. B. (eds.) Bacteriophage T4, American Society for Microbiology, pp. 174-183.

Burnett, R. M., Darling, G. D., Kendall, D. S., LeQuesne, M. E., Mayhew, S. G., Smith, W. W, and Ludwig, M. L. (1974) J.Biol. Chem., 249, 4383-4392.

Cameron, V., Soltis, D., and Uhlenbeck, O. C. (1978) Nucleic Acids Res. 5, 825-833.

Cameron, V. and Uhlenbeck, O. C. (1977) Biochemistry, 16, 5120-5126.

Carlson, K., and Nicolaisen, B. (1979) J.Virol. 31, 112-123.

Cascino, A., Riva, S. and Geiduscheck, E. P. (1970) Cold Spring Harbor Symp. Quant. Biol., 35, 213-220.

Casjens, S. and King, J. (1975) Annu.Rev.Biochem. 44, 555-611.

Chan, V. L. and Ebisuzaki, K. (1970) Mol.Gen.Genet. 109, 162-168.

Christensen, A. C. and Young, E. T. (1982) Nature (London) 299, 369-371.

Chu, F. K., Maley, G. F., Belfort, M. and Maley, F. (1985) J.Biol.Chem. in press

Chu, F. K., Maley, G. F., Maley, F. and Belfort, M. (1984) Proc.Natl.Acad.Sci.USA, 81, 3049-3053.

Cohen, S. A., Sterner, R., Keim, P. S. and Heinrikson, R. L. (1978) J.Biol.Chem., 253, 889-897.

Cohn, M., Leigh, J. S. and Reed, G. H. (1971) Cold Spring Harbor Symp. Quant. Biol., 36, 533-540.

Cooley, W. C., Sirotkin, K., Green, R. and Snyder, L. (1979) J.Bacteriol. 140, 83-91.

Daegelen, P., D'Aubeuton-Carafa, Y. and Brody, E. (1982) Virology, 117, 121-134.

Dannenberg, R. and Mosig, G. (1983) J.Virol. 45, 813-831.

David, M., Borasio, G. D. and Kaufmann, G. (1982) Proc. Natl.Acad.Sci.USA, 79, 7097-7101.

David, M., Veckstein, R. and Kaufmann, G. (1979) Proc.Natl.Acad.Sci.USA, 76, 5430-5434.

Davis, B. D. and Tai, P. C. (1980) Nature (London), 283, 433-438.

Delius, H., Howe, C. and Kozinski, A. W. (1971) Proc.Natl.Acad.Sci.USA, 68, 3049-3053.

Depew, R. E. and Cozzarelli, N. R. (1974) J.Virol. 13, 888-897.

Dube, S. K. and Rudland, P. S. (1970) Nature (London), 226, 820-823.

Edgar, R. S., Denhardt, G. H. and Epstein, R. T. (1964) Genetics, 49, 635-648.

Edgar, R. S. and Wood, W. B. (1966) Proc.Natl.Acad.Sci. USA, 55, 498-505.

Elliot, T. and Geiduschek, E. P. (1984) Cell, 36, 211-218.

Ellis, E. L. and Delbruck, M. (1939) J.Gen.Physiol. 22, 365-384.

Emrich, J. (1968) Virology, 35, 158-165.

Epstein, R. H., Bolle, A., Steinberg, C., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R., Susman, M., Denhardt, C. and Lielausis, I. (1963) Cold Spring Harbor Symp. Quant. Biol. 28, 375-392.

Folkmanis, A., Maltzman, W., Mellon, P., Skalka, A. and Echols, H. (1977) Virology, 81, 352-362.

Franklin, N. C. (1971) In Hershey, A. D. (ed.) The Bacteriophage Lambda, Cold Spring Harbor Laboratory,pp 175-194

Franklin, N. C. (1974) J.Mol.Biol., 89, 33-48.

Friedman, D. and Gottesman, M. (1983) In Hendrix, R. W., Roberts, J. W., Stahl, F. W., Weisberg, R. A. (eds.) Lambda II, Cold Spring Harbor Laboratory, pp 21-51.

Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. E. (1983) J.Mol.Biol., 170, 827-842.

Georgopoulos, C. P. (1967) Biochem.Biophys.Res.Commun., 28, 179-184.

Goff, C. G. (1974) J.Biol.Chem., 249, 6181-6190.

Goff, C. G. and Weber, K. (1970) Cold Spring Harbor Symp. Quant. Biol. 35, 101-108.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. and Stormo, G. (1981) Annu.Rev.Microbiol. 35, 365-403.

Goldberg, A. R. (1970) Cold Spring Harbor Symp. Quant. Biol., 35, 157-161.

Goldfarb, A. (1981a) Nucleic Acids Res. 9, 519-527.

Goldfarb, A. (1981b) Proc. Natl.Acad.Sci.USA 78, 3454-3458.

Goodman, H. M., Olson, M. V. and Hall, B. D. (1977) Proc.Natl.Acad.Sci.USA, 74, 5453-5457.

Gough, J. A. and Murray, N. E. (1983) J.Mol.Biol., 166, 1-19.

Goulian, M., Lucas, Z. J. and Kornberg, A. (1968) J.Biol.Chem. 243, 627-638.

Gram, H., Liebig, H.-D., Hack, A., Niggemann, E. and Ruger, W. (1984) Mol.Gen.Genet., 194, 232-240.

Greenblatt, J. (1981) Cell, 24, 8-9.

Greer, C. L., Javor, B., and Abelson, J. (1983) Cell, 33, 899-906.

Guarente, L., Lauer, G., Roberts, T. M. and Ptashne, M. (1980) Cell, 20, 543-553.

Guha, A., Szybalski, W., Salser, W., Bolle, A., Geiduschek, E. P. and Pulitzer, J. F. (1971) J.Mol.Biol., 59, 329-349.

Guthrie, C, and McClain, W. H. (1973) J.Mol.Biol. 81, 135-137.

Halpern, M. E., Mattson, T. and Kozinski, A. W. (1979) Proc.Natl.Acad.Sci.USA, 76, 6137-6141.

Hamilton, S. and Pettijohn, D. E. (1976) J.Virol. 19, 1012-1027.

Hanahan, D. (1983) J.Mol.Biol., 166, 557-580.

Hancock, R. E. W., Hankte, K. and Braun, V. (1976) J.Bacteriol. 127, 1370-1375.

Haselkorn, R., Vogel, M. and Brown, R. D. (1969) Nature (London), 221, 836-838.

Hattman, S. (1970) Virology, 42, 359-367.

Hattman, S. and Fukasawa, T. (1963) Proc.Natl.Acad.Sci.USA, 50, 297-300.

Hershey, A. D. and Chase, M. (1952), J.Gen.Physiol. 36, 39-56.

Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. and Helinski, D. R. (1974) Proc.Natl.Acad.Sci.USA, 71, 3455-3459.

Horvitz, H. R. (1973) Nature (London) New Biol. 244, 137-140.

Howe, C. C., Buckley, P. J., Carlson, K. M. and Kozinski, A. W. (1973) J.Virol. 12, 130-148.

Hsu, W.-T. and Weiss, S. B. (1969) Proc.Natl.Acad.Sci.USA, 64, 345-351.

Huang, C. C., Hearst, J. E. and Alberts, B. M. (1981) J.Biol.Chem. 256, 4087-4094.

Hughes, S. G. and Brown, P. R. (1973) Biochem.J., 131, 583.

Jaskunas, S. R., Lindahl, L., Nomura, M. and Burgess, R. R. (1975) Nature (London), 257, 458-462.

Jay, G., Khoury, G., Seth, A. K. and Jay, E. (1981) Proc.Natl.Acad.Sci.USA, 78, 5543-5548.

Jayaraman, R. (1972) J.Mol.Biol. 70, 253-263.

Jayaraman, R. and Goldberg, E. B. (1970) Cold Spring Harbor Symp.Quant.Biol., 35, 197-201.

Jencks, W. P. (1975) Advan.Enzymol. 43, 219-410.

Josslin, R. (1970) Virology, 40, 719-726.

Josslin, R. (1971) Virology, 44, 101-107.

Kaplan, D. A. and Nierlich, D. P. (1975) J.Biol.Chem. 250, 2395-2397.

Karam, J. D., Gold, L., Singer, B. S. and Dawson, M. (1981) Proc.Natl.Acad.Sci.USA, 78, 4669-4673.

Kassavetis, G. A. and Geiduschek, E. P. (1982) EMBO J. 1. 107-114.

Kassavetis, G. A. and Geiduschek, E. P. (1984) Proc.Natl. Acad.Sci.USA, 81, 5101-5105.

Kelley, W. S., Chalmers, K. and Murray, N. E. (1977) Proc.Natl.Acad.Sci.USA, 74, 5632-5636.

Kemper, B. and Brown, D. T. (1976) J.Virol., 18, 1000-1015.

Kemper, B., Garabath, M. and Courage, V (1981) Eur.J. Biochem. 115, 133-141.

Kennell, D. (1970) J.Virol., 6, 208-217.

Khesin, R. B., Bogdanova, E. S., Goldfarb, A. D. and Zograff, Y. N. (1972) Mol.Gen.Genet. 119, 299-314.

Khesin, R. B., Nikiforov, V. G., Zograff, Y. N., Danilevskaya, D. N., Kalayaeva, E. S., Lipkin, V. M., Modyanov, N. N., Dmitriev, A. D., Velkov, V. V. and Gintsburg, A. L. (1976) In Losick, R. and Chamberlin, M. (eds.) RNA Polymerase, Cold Spring Harbor Laboratory pp. 629-643.

Khorana, H. G., Agarwal, K. D., Büchi, H., Caruthers, M. H., Gupta, N. K., Kleppe, K., Kumar, A., Ohtsuka, E., Raj Bhandary, U. L., Van de Sande, J. H., Sgaramella, V., Terao, T., Weber, H. and Yamada, T. (1972) J.Mol.Biol., 72, 209-217.

Kiko, H., Niggemann, E. and Ruger, W. (1979) Mol.Gen. Genet., 172, 303-312.

Kim, J. S. and Davidson, N. (1974) Virology, 57, 93-111.

Kleppe, K., and Lillehaug, J. R. (1979) Adv.Enzymol. 38, 245-275.

Konarska, M., Fillipowicz, W. and Gross, H. J. (1982) 79, 1474-1478.

Kornberg, S. R., Zimmerman, S. B. and Kornberg, A. (1961) J.Biol.Chem. 236, 1487-1493.

Koshland, D. E. Jr. (1958) Proc.Natl.Acad.Sci. USA, 44 98-104.

Kozinski, A. W. (1961) Virology, 13, 124-128.

Kozinski, A. W. and Doermann, A. H. (1975) Proc.Nat.Acad. Sci.USA, 72, 1734-1738.

Knapp, G., Ogden, R. C., Peebles, C. L. and Abelson, J. (1977) Cell, 18, 37-45.

Krisch, H. M., Bolle, A. and Epstein, R. (1974) J.Mol. Biol., 88, 89-104.

Kutter, E., Drivdahl, R., and Rand, K. (1984) Genetics, 108, 291-304.

Kutter, E., O'Farrell, P. and Guttman, B. (1980) In O'Brien, S. (ed.) N.I.H. Genetic Maps, vol. I. National Institutes of Health, pp. 33-40.

Kutter, E. and Rüger, W. (1983) In Mathews, C., Kutter, E. M., Mosig, G. and Berget, P. E. (eds.) Bacteriophage T4, American Society for Microbiology, pp. 277-290.

Kutter, E. and Wiberg, J. (1969) J.Virol., 4, 439-453.

Lederberg, E. M. and Cohen, S. N. (1974) J.Bacteriol. 119, 1072-1074.

Lehman, I. R. and Nussbaum, A. L. (1968) J.Biol.Chem., 239, 2628-2636.

Lehman, I. R. and Pratt, E. A. (1960) J.Biol.Chem. 235, 3254-3258

Lembach, K. J., Kuninaka, A. and Buchanan, J. M. (1969) Proc.Natl.Acad.Sci.USA, 62, 446-453.

Li, L., Tanyashin, V. I., Matvienko, N. I. and Bayev, A. A. (1975) Dokl.Acad.Nauk. SSSR., 223, 1262-1265.

Lillehaug, J. R. (1977) Eur.J.Biochem., 73, 499-506.

Lillehaug, J. R. and Kleppe, K. (1975a) Biochemistry, 14, 1221-1225.

Lillehaug, J. R. and Kleppe, K. (1975b) Biochemistry 14, 1225-1229.

Lillehaug, J. R., Kleppe, R. K. and Kleppe, K. (1976) Biochemistry, 15, 1858-1864.

Liu, C. C. and Alberts, B. M. (1981) J.Biol.Chem., 256, 2813-2820.

Lowry, O. H., Rosebrough, N. J., Farr, D. L. and Randall, R. J. (1951) J.Biol.Chem., 193, 265-275.

Luder, A. and Mosig, G. (1982) Proc.Natl.Acad.Sci.USA, 79, 1101-1105.

Luftig, R. B., Wood, W. B. and Okinaka, R. (1971) J.Mol.Biol. 57, 555-573.

Mace, D. C. and Alberts, B. M. (1984) J.Mol.Biol. 177, 279-293.

Mailhammer, R., Yang, H.-L., Reiness, G. and Zubay, G. (1975) Proc.Natl.Acad.Sci.USA, 72, 4828-4932.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory.

Matsubara, K. (1976) Am.Soc.Microbiol. News, 42, 134.

Mattson, T., van Howe, G., Bolle, A., Selzer, G. and Epstein, R. (1977) Mol.Gen.Genet. 154, 319-326.

Mattson, T., van Howe, G., and Epstein, R. H. (1978) J.Mol.Biol. 126, 551-570.

Mattson, T., Richardson, J. and Goodin, D. (1974) Nature (London) 250, 48-50.

Maxam, A. M. and Gilbert, W. (1977) Proc.Natl.Acad.Sci.USA, 74, 560-564.

McClain, W. H., Guthrie, C. and Barrell, (1972) Proc.Natl. Acad.Sci.USA, 69, 3703-3707.

McKenney, K., Shimitake, H., Court, D., Schmeissner, U., Brody, C. and Rosenberg, M. (1981) In Chirikjian, J. G. and Papas, T. S. (eds.) Gene Amplification and Analysis vol. II: Structural Analysis of Nucleic Acids, Elsevier, pp. 383-415.

Messing, J. (1983) Methods in Enzymology, 101, 20-78.

Mileham, A. J., Revel, H. R. and Murray, N. E. (1980) Mol.Gen.Genet., 179. 227-239.

Mizuuchi, K., Kemper, B., Hays, J., and Weisberg, R. A. (1982) Cell, 29, 357-365.

Moir, A. and Brammar, W. J. (1976) Mol.Gen.Genet., 149, 87-99.

Monod, J. and Wollman, E. L. (1947) Ann.Inst.Pasteur Paris, 73, 937-956.

Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. and Rossman, M. G. (1975) J.Biol.Chem. 250, 9137-9162.

Morton, D., Kutter, E. M. and Guttman, B. S. (1975) J. Virol. 28, 262-269.

Mosig, G. (1968) Genetics, 59, 137-151.

Mosig, G. (1983) In Mathews, C. K., Kutter, E. M., Mosig, G. and Berget, P. B. (eds.) Bacteriophage T4, American Society for Microbiology, pp 362-374.

Mosig, G. D., Ghosal, D. and Bock, S. (1981) In Dubow, M. (ed.) Bacteriophage Assembly, Allen R. Liss Inc. pp 139-150.

Müller-Hill, B., Crapo, L. and Gilbert, W. (1968) Proc.Natl. Acad. Sci.USA, 59, 1259-1264.

Muller, U. H. and Marchin, G. L. (1975) J.Virol. 15, 238-243.

Murray, K. (1970) Biochem. J., 118, 831-841.

Murray, K. (1973) Biochem. J., 131, 569-583.

Murray, N. E. (1983) In Hendrix, R. W., Roberts, J. W., Stahl, F. W. and Weisberg, R. A. (eds.) Lambda II, Cold Spring Harbor Laboratory, pp. 395-432.

Murray, N. E., Brammar, W. J. and Murray, K. (1977) Mol.Gen.Genet. 150, 53-61.

Murray, N. E., Bruce, S. A. and Murray, K. (1979) J.Mol.Biol. 132, 493-505.

Murray, N. E. and Kelley, W. S. (1979) Mol.Gen.Genet., 175, 77-87.

Nakamura, K. and Inouye, M. (1982) EMBO J, 1, 771-775.

Nelson, M. A., Singer, B. S., Gold, L. and Pribnow, D. (1981) J.Mol.Biol. 149, 377-403.

Noda, L. (1958) J.Biol.Chem. 232, 237-247.

Novogrodsky, A. and Hurwitz, J. (1966) J.Biol.Chem. 241, 2923-2932.

Novogrodsky, A., Tal, M., Traub, A. and Hurwitz, J. (1966) J. Biol.Chem. 241, 2933-2943.

O'Farrell, P. H., Kutter, E. and Nakanish, M. (1980) Mol.Gen.Genet. 179, 421-435.

O'Farrell, P. Z. and Gold, L. M. (1973) J.Biol.Chem. 248, 5502-5511.

O'Farrell, P. Z., Gold, L. M. and Huang, W. M. (1973) J.Biol.Chem. 248, 5499-5505.

Pai, E. F., Sachsenheimer, W., Schirmer, R. H. and Schulz, G. E. (1977) J.Mol.Biol. 114, 37-45.

Panasenko, S. N., Cameron, J. R., Davis, R. W. and Lehman, I. R. (1977) Science, 196, 188-189.

Panet, A., van de Sande, J. H., Loewen, P. C., Khorana, H. G., Raae, A. J., Lillehaug, J. R. and Kleppe, K. (1973) Biochemistry, 12, 5045-5049.

Piperno, J. R., Kallen, R. B. and Alberts, B. M. (1978) J.Biol.Chem., 253, 5180-5185.

Pribnow, D., Sigurdson, D. C., Gold, L., Singer, B. S., Napoli, C., Brosius, J., Dull, T. J. and Noller, H. F. (1981) J.Mol.Biol. 149, 337-376.

Ptashne, M. (1971) In Hershey, A. D. (ed.) The Bacteriophage Lambda, Cold Spring Harbor Laboratory, pp 221-237.

Rabussay, D. (1982) ASM News, 48, 398-403.

Rabussay, D. and Geiduschek, E. P. (1977a) Compr.Virol., 8, 1-196.

Rabussay, D. and Geiduschek, E. P. (1977b) Proc.Natl.Acad. Sci.USA, 74, 5305-5309.

Rabussay, D., Mailhammer, R. and Zillig, W. (1972) In Wieland, O., Helmreich, E. and Holzer, H. (eds.) Metabolic Interconversion of Enzymes, Springer-Verlag, pp 213-227.

Rand, K. N. and Gait, M. J. (1984) EMBO J, 3, 397-402.

Ratner, D. (1974) J.Mol.Biol. 88, 373-383.

Remaut, E., Stanssens, P. and Fiers, W. (1981) Gene, 15, 81-93.

Remaut, E., Tsao, H. and Fiers, W. (1983) Gene, 22, 103-113.

Revel, H. R. and Luria, S. (1970) Annu.Rev.Genet., 4, 177-192.

Rhoads, D. G. and Lowenstein, J. M. (1968) J.Biol.Chem. 243, 3963-3972.

Richardson, C. C. (1965) Proc.Natl.Acad.Sci.USA, 54, 158-165.

Richardson, J. P. (1970) Cold Spring Harbor Symp.Quant. Biol. 35, 127-133.

Ritchie, D. A. and White, F. E. (1972) J.Gen.Virol., 16, 91-94.

Riva, S., Cascino, A. and Geiduschek, E. P. (1970a) J.Mol.Biol. 54, 85-102.

Riva, S., Cascino, A. and Geiduschek, E. P. (1970b) J.Mol.Biol. 54, 103-119.

Roberts, J. W. (1970) Cold Spring Harbor Symp. Quant. Biol. 35, 121-126.

Rohrer, H., Zillig, W. and Mailhammer, R. (1975) Eur.J. Biochem. 60, 227-238.

Rosenberg, M., and Court, D. (1979) Annu.Rev.Genet. 13, 319-353.

Rüger, W., Newmann, M., Rohr, U. and Niggeman, E. (1979) Mol.Gen.Genet. 176, 417-425.

Runnels, J. M., Soltis, D., Hey, T. and Snyder, L. (1982) J.Mol.Biol. 154, 273-286.

Sachsenheimer, W. and Schulz, G. E. (1977) J.Mol.Biol. 114, 23-36.

Sadowski, P. D. and Bakyta, I. (1972) J.Biol.Chem. 247, 405-412.

Sadowski, P. D. and Hurwitz, J. (1969) J.Biol.Chem. 244, 6182-6191.

Salser, W., Bolle, A. and Epstein, R. (1970) J. Mol.Biol. 49, 271-295.

Sanger, F., Brownlee, G. G. and Barrell, B. G. (1965) J.Mol.Biol. 13, 373-398.

Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. and Peterson, G. B. (1982) J.Mol.Biol. 162, 729-773.

Sano, H. (1976) Biochim.Biophys.Acta, 422, 109-119.

Schachner, M., Seifert, W. and Zillig, W. (1971) Eur.J. Biochem. 22, 520-528.

Schulz, G. E. and Schirmer, R. H. (1974) Nature (London) 250, 142-144.

Secrist, J. A., Barrio, J. R., Leonard, N. J. and Weber, G. (1972) Biochemistry, 11, 3499-3506.

Selzer, G., Bolle, A., Krisch, H. and Epstein, R. (1978) Mol.Gen.Genet. 159, 301-309.

Shimitake, H. and Rosenberg, M. (1981) Nature (London) 292, 128-132.

Shine, J. and Dalgarno, L. (1974) Proc.Natl.Acad.Sci.USA, 71, 1342-1346.

Siebenlist, U., Simpson, R. B. and Gilbert, W. (1980) Cell, 20, 269-281.

Silver, L. L. and Nossal, N. G. (1982) J.Biol.Chem., 257, 11696-11705.

Silverstein, J. L. and Goldberg, E. B. (1976) Virology, 72, 212-223.

Simon, L. D. and Anderson, T. F. (1967) Virology, 32, 279-297.

Sirotkin, L., Cooley, W., Runnels, J., and Snyder, L. R. (1978) J.Mol.Biol. 123, 221-233.

Slotin, L. A. and Hampton, A. (1975) Biochemistry, 14, 4538-5446.

Sly, W. S., Rabideau, K. and Kolber, A. (1971) In Hershey, A. D. (ed.) The Bacberiophage Lambda, Cold Spring Harbor Laboratory, pp. 575-588.

Snopek, T. J., Wood, W. B., Conley, M. P., Chen, P. and Cozzarelli, N. R. (1977) Proc.Natl.Acad.Sci.USA, 74, 3355-3359.

Snyder, L. (1983) In Mathews, C. K., Kutter, E. M., Mosig, G. and Berget, P. B. Bacteriophage T4, American Society for Microbiology, pp 351-355.

Snyder, L., Gold, L. and Kutler, E. (1976) Proc.Natl. Acad.Sci.USA, 73, 3098-3102.

Staden, R. (1982) Nucleic Acids Res., 10, 4731-4751.

Stahl, F. W., Edgar, R. S. and Steinberg, J. (1964) Genetics, 50, 539-552.

Stahl, F. W. and Murray, N. E. (1966) Genetics, 53, 569-576.

Stevens, A. (1972) Proc.Natl.Acad.Sci.USA, 69, 603-607.

Stevens, A. (1977) Biochim.Biophys.Acta, 475, 193-196.

Streisinger, G., Edgar, R. S. and Denhardt, G. H. (1964) Proc.Nat.Acad.Sci.USA, 51, 775-779.

Takahashi, H., Shimizu, M., Saito, H. and Ikeda, Y. (1979) Mol.Gen.Genet. 168, 49-53.

Thomas, C. A. Jr. and Rubenstein, I. (1964) Biophys.J. 4, 93-106.

Trimble, R. B. and Maley, F. (1976) J.Virol. 17, 538-549.

Tsugita, A. (1971) In Boyer, P. D. (ed.) The Enzymes, Vol. V, Academic Press, pp. 343-411.

Uhlenbeck, O. C. and Gumport, R. I. (1982) In Boyer, P. D. (ed.) The Enzymes, Vol. XV, Academic Press, pp. 31-58.

Uhlin, B. E., Molin, S. Gustafsson, P. and Nordstrom, K. (1979) Gene 6, 91-106.

Vanderslice, R. W. and Yegian, C. D. (1974) Virology, 60, 265-275.

van de Sande, J. H., Kleppe, K. and Khorana, H. G. (1973) Biochemistry, 12, 5050-5055.

Velten, J. and Abelson, A. (1980) J.Mol.Biol. 137, 235-248.

Velten, J., Fukada, K. and Abelson, J. (1976) Gene 1, 93-106.

von Zabern, I., Wittman-Liebold, B., Untucht-Grau, R., Schirmer, R.H. and Pai, E. F. (1976) Eur.J.Biochem. 68, 281-290.

Vorozheikina, D., Glinskaite, I., Tikhomirova, L. and Bayev, A. (1980) Mol.Gen.Genet., 178, 655-661.

Walker, J. E.,Saraste, M., Runswick, M. J. and Gray, N. J. (1982) EMBO J, 1, 945-951.

Warner, H. R., Snustad, D. P., Jorgensen, S. E. and Koerner, J. F. (1970) J.Virol. 5, 700-708.

Weigle, J. (1966) Proc.Natl.Acad.Sci.USA, 55, 1462-1466.

Weil, J., Cunningham, R., Martin, R., Mitchell, B. and Bolling, B. (1973) Virology, 50, 373-380.

Wiberg, J. S., Mendelsohn, S., Warner, V., Hercules, K., Aldrich, C. and Munro, L. (1973) J.Virol. 12, 1755-1792.

Wilson, G. G. and Murray, N. E. (1979) J.Mol.Biol., 132, 471-491.

Wilson, G. G., Tanyashin, V. I. and Murray, N. E. (1977) Mol.Gen.Genet., 156, 203-214.

Wilson, J. H., Luftig, R. B. and Wood, W. (1970) J.Mol. Biol., 51, 423-434.

Wood, W. B. and Revel, H. R. (1976) Bacteriol. Rev. 40, 847-868.

Wu, A. M., Ghosh, S., Echols, H. and Spiegelman, W.G. (1972) J.Mol.Biol. 67, 407-421.

Wu, R., Geiduschek, E. P. and Cascino, A. (1975) J. Mol. Biol. 96, 539-562,

Wyatt, G. R. and Cohen, S. S. (1953) Biochem.J. 55, 774-782.

Yanisch-Perron, C., Viera, J. and Messing, J. (1985) Gene, 33, 103-119.

Young, E. T., (1975) J.Mol.Biol. 96, 393-424.

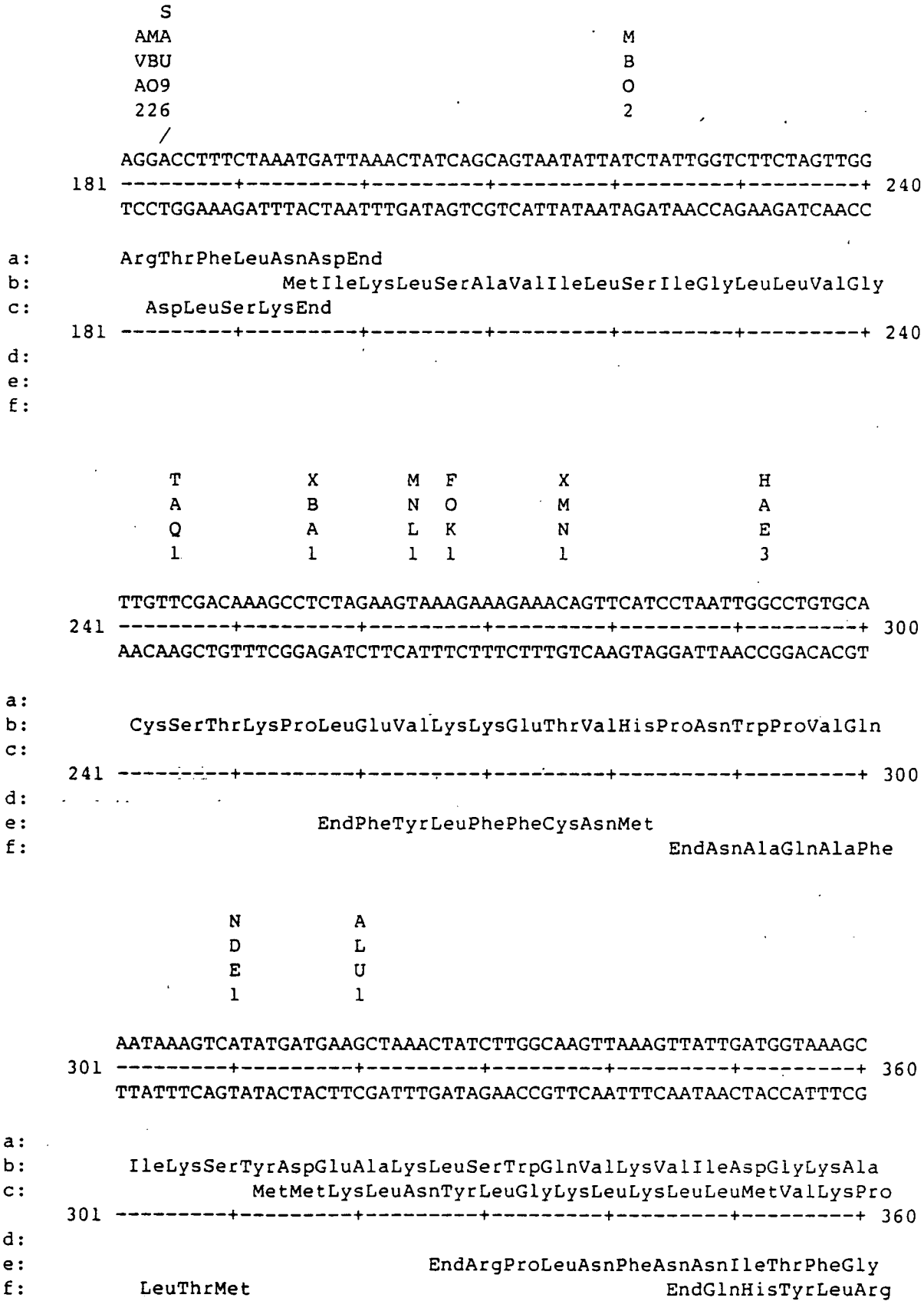
Young, E. T., Mattson, T., Selzer, G., van Houwe, G., Bolle, A. and Epstein, R. (1980) J.Mol.Biol. 138, 423-445.

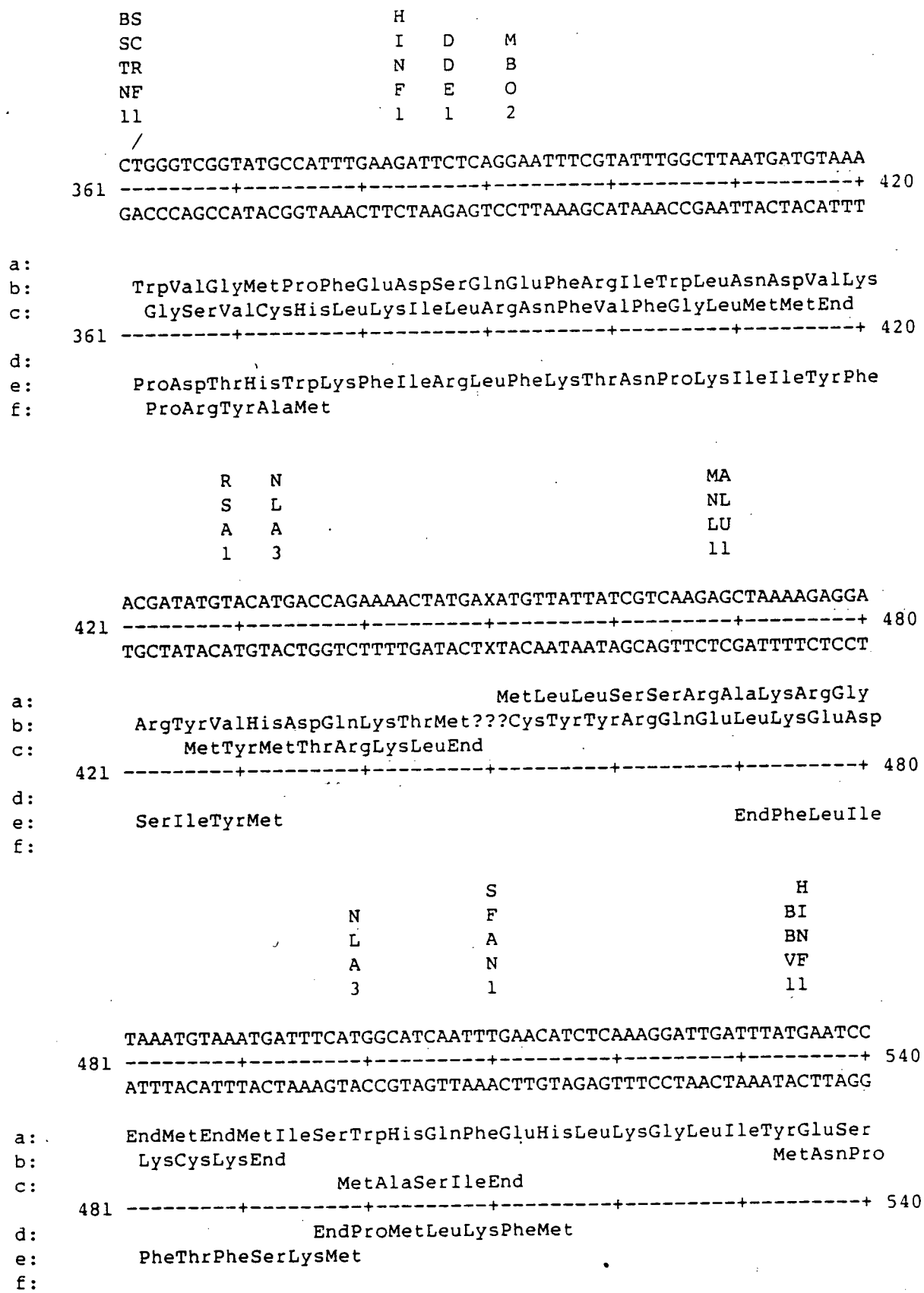
Yudelevich, A. (1971) J.Mol.Biol. 60, 21-29.

APPENDIX I

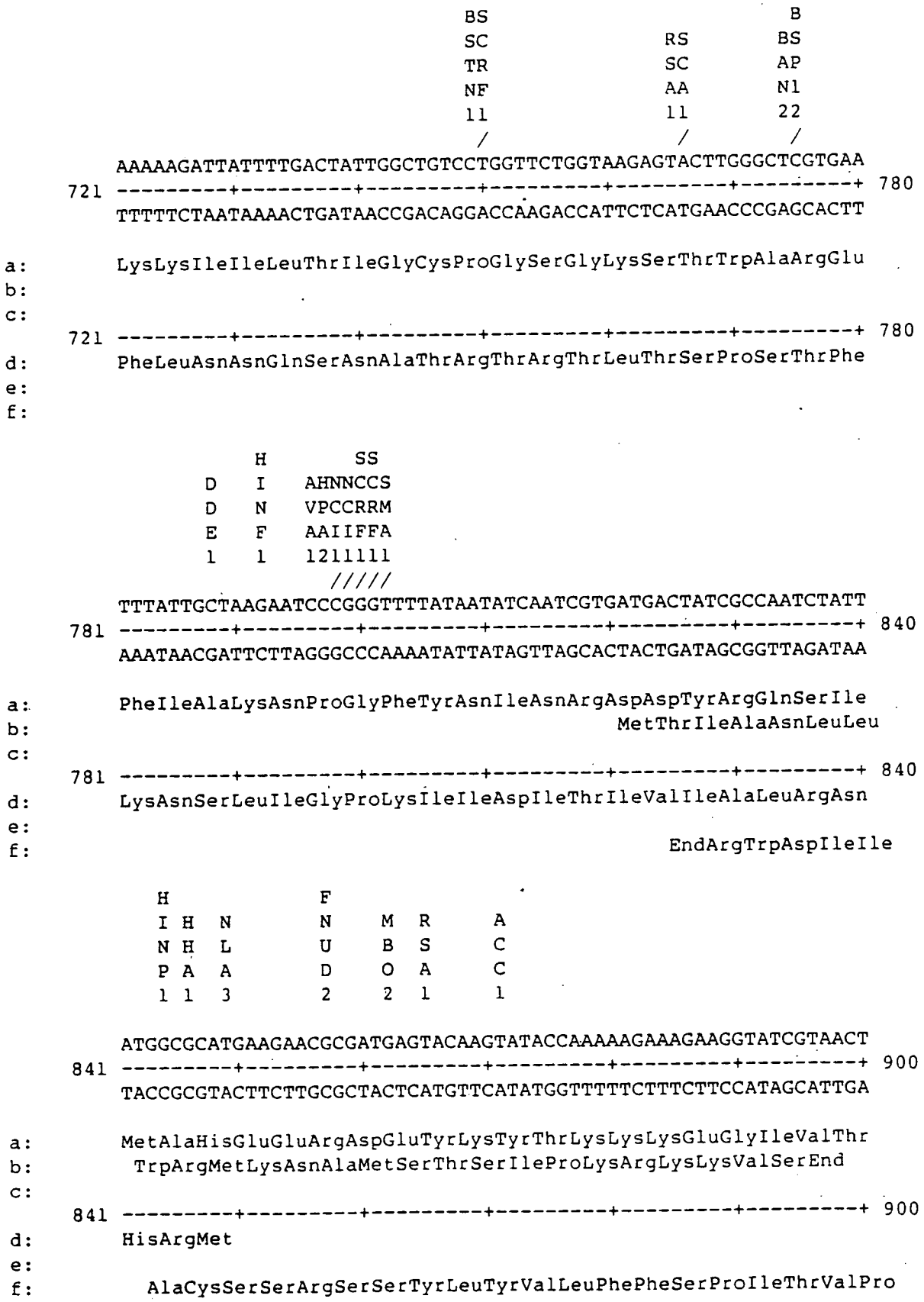
The DNA sequence of the contiguous T4 EcoRI fragments 21 and 46 (Figure 3.1). The open reading frames on both strands are translated below, and restriction enzyme target sites are shown above the T4 DNA sequence.

```
    E         S                                    M   B            F
    C         F                                    N   B            N
    O         A                                    N   B            U
    R         N                                    L   V            4
    1         1                                    1   1            H

    GAATTCAGTATAATTATATTGATGCGATGAATAATAAAAATCGTGAGGCAATTGCTGCTA
  1 ---------+---------+---------+---------+---------+---------+ 60
    CTTAAGTCATATTAATATAACTACGCTACTTATTATTTTTAGCACTCCGTTAACGACGAT

a:  GluPheSerIleIleIleLeuMetArgEnd
b:   AsnSerValEnd
c:    IleGlnTyrAsnTyrIleAspAlaMetAsnAsnLysAsnArgGluAlaIleAlaAlaIle
  1 ---------+---------+---------+---------+---------+---------+ 60
d:
e:
f:    AsnLeuIleIleIleAsnIleArgHisIleIlePheIleThrLeuCysAsnSerSerAsn

                    S  H
                    F  IMH                                    F
                    A  NSH                                    O
                    N  PTA                                    K
                    1  111                                    1

    TTGAGCGTGAAAATGAAAAACTGCGCAAAGATGCAAAGAAGGCGGATGTGGTGGCTCATA
 61 ---------+---------+---------+---------+---------+---------+ 120
    AACTCGCACTTTTACTTTTTGACGCGTTTCTACGTTTCTTCCGCCTACACCACCGAGTAT

a:            MetLysAsnCysAlaLysMetGlnArgArgArgMetTrpTrpLeuIle
b:
c:    GluArgGluAsnGluLysLeuArgLysAspAlaLysLysAlaAspValValAlaHisLys
 61 ---------+---------+---------+---------+---------+---------+ 120
d:
e:                                                           EndLeu
f:    LeuThrPheIlePhePheGlnAlaPheIleCysLeuLeuArgIleHisHisSerMet

       BS                                                    BS
       SC                                   F                SC
       TR                                   O                TR
       NF                                   K                NF
       11                                   1                11
        /                                                     /
    AGCCAGGATTGGTTGAAAAACAAATCAACAACTCCTTCAACAAGTTCGCAGAAGACATCC
121 ---------+---------+---------+---------+---------+---------+ 180
    TCGGTCCTAACCAACTTTTTGTTTAGTTGTTGAGGAAGTTGTTCAAGCGTCTTCTGTAGG

a:  SerGlnAspTrpLeuLysAsnLysSerThrThrProSerThrSerSerGlnLysThrSer
b:
c:    ProGlyLeuValGluLysGlnIleAsnAsnSerPheAsnLysPheAlaGluAspIleGln
121 ---------+---------+---------+---------+---------+---------+ 180
d:
e:   GlyProAsnThrSerPheCysIleLeuLeuGluLysLeuLeuAsnAlaSerSerMet
f:
```

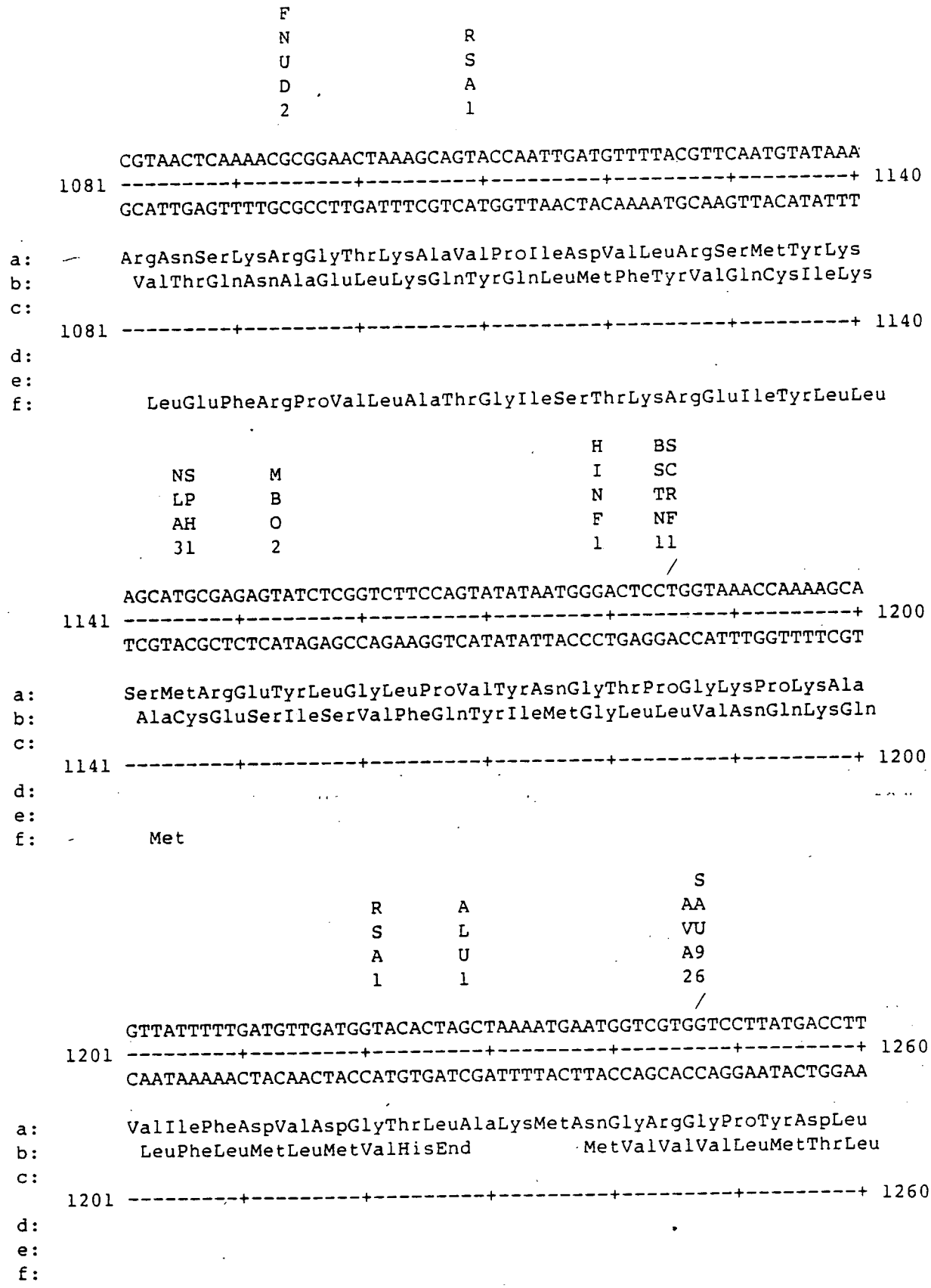

S
AMA
VBU
AO9
226
/
M
B
O
2
AGGACCTTTCTAAATGATTAAACTATCAGCAGTAATATTATCTATTGGTCTTCTAGTTGG
181 ---------+---------+---------+---------+---------+---------+ 240
TCCTGGAAAGATTTACTAATTTGATAGTCGTCATTATAATAGATAACCAGAAGATCAACC
a: ArgThrPheLeuAsnAspEnd
b: MetIleLysLeuSerAlaValIleLeuSerIleGlyLeuLeuValGly
c: AspLeuSerLysEnd
181 ---------+---------+---------+---------+---------+---------+ 240
d:
e:
f:
T
A
Q
1
X
B
A
1
M
N
L
1
F
O
K
1
X
M
N
1
H
A
E
3
TTGTTCGACAAAGCCTCTAGAAGTAAAGAAAGAAACAGTTCATCCTAATTGGCCTGTGCA
241 ---------+---------+---------+---------+---------+---------+ 300
AACAAGCTGTTTCGGAGATCTTCATTTCTTTCTTTGTCAAGTAGGATTAACCGGACACGT
a:
b: CysSerThrLysProLeuGluValLysLysGluThrValHisProAsnTrpProValGln
c:
241 ---------+---------+---------+---------+---------+---------+ 300
d:
e: EndPheTyrLeuPhePheCysAsnMet
f: EndAsnAlaGlnAlaPhe
N
D
E
1
A
L
U
1
AATAAAGTCATATGATGAAGCTAAACTATCTTGGCAAGTTAAAGTTATTGATGGTAAAGC
301 ---------+---------+---------+---------+---------+---------+ 360
TTATTTCAGTATACTACTTCGATTTGATAGAACCGTTCAATTTCAATAACTACCATTTCG
a:
b: IleLysSerTyrAspGluAlaLysLeuSerTrpGlnValLysValIleAspGlyLysAla
c: MetMetLysLeuAsnTyrLeuGlyLysLeuLysLeuLeuMetValLysPro
301 ---------+---------+---------+---------+---------+---------+ 360
d:
e: EndArgProLeuAsnPheAsnAsnIleThrPheGly
f: LeuThrMet EndGlnHisTyrLeuArg


BS H
SC I D M
TR N D B
NF F E O
11 1 1 2
CTGGGTCGGTATGCCATTTGAAGATTCTCAGGAATTTCGTATTTGGCTTAATGATGTAAA
361
GACCCAGCCATACGGTAAACTTCTAAGAGTCCTTAAAGCATAAACCGAATTACTACATTT
420
a:
b: TrpValGlyMetProPheGluAspSerGlnGluPheArgIleTrpLeuAsnAspValLys
c: GlySerValCysHisLeuLysIleLeuArgAsnPheValPheGlyLeuMetMetEnd
361
420
d:
e: ProAspThrHisTrpLysPheIleArgLeuPheLysThrAsnProLysIleIleTyrPhe
f: ProArgTyrAlaMet
R N MA
S L NL
A A LU
1 3 11
ACGATATGTACATGACCAGAAAACTATGAXATGTTATTATCGTCAAGAGCTAAAAGAGGA
421
TGCTATACATGTACTGGTCTTTTGATACTXTACAATAATAGCAGTTCTCGATTTTCTCCT
480
a: MetLeuLeuSerSerArgAlaLysArgGly
b: ArgTyrValHisAspGlnLysThrMet???CysTyrTyrArgGlnGluLeuLysGluAsp
c: MetTyrMetThrArgLysLeuEnd
421
480
d:
e: SerIleTyrMet EndPheLeuIle
f:
S H
N F BI
L A BN
A N VF
3 1 11
TAAATGTAAATGATTTCATGGCATCAATTTGAACATCTCAAAGGATTGATTTATGAATCC
481
ATTTACATTTACTAAAGTACCGTAGTTAAACTTGTAGAGTTTCCTAACTAAATACTTAGG
540
a: EndMetEndMetIleSerTrpHisGlnPheGluHisLeuLysGlyLeuIleTyrGluSer
b: LysCysLysEnd MetAsnPro
c: MetAlaSerIleEnd
481
540
d: EndProMetLeuLysPheMet
e: PheThrPheSerLysMet
f:

```
                F               H   H
                N               A     I H          I
                U               H     N G          N
                4               A     F A          F
                H               2     1 1          1

         GAGATGGCTGCAATGATTTATGGACGCCAGATTCAGCGGTTAGAATCTTTACCTCCAACT
     541 ---------+---------+---------+---------+---------+---------+ 600
         CTCTACCGACGTTACTAAATACCTGCGGTCTAAGTCGCCAATCTTAGAAATGGAGGTTGA

a:       GluMetAlaAlaMetIleTyrGlyArgGlnIleGlnArgLeuGluSerLeuProProThr
b:        ArgTrpLeuGlnEnd   MetAspAlaArgPheSerGlyEnd
c:
     541 ---------+---------+---------+---------+---------+---------+ 600
d:
e:
f:


          M             A
          N             L
          L             U
          1             1

         AATGATGTTTTATTAGCTCAATCACGTGCTAATCTCAAAAATGAATATCAAAATAAGTGG
     601 ---------+---------+---------+---------+---------+---------+ 660
         TTACTACAAAATAATCGAGTTAGTGCACGATTAGAGTTTTTACTTATAGTTTTATTCACC

a:       AsnAspValLeuLeuAlaGlnSerArgAlaAsnLeuLysAsnGluTyrGlnAsnLysTrp
b:        MetMetPheTyrEnd                        MetAsnIleLysIleSerGly
c:
     601 ---------+---------+---------+---------+---------+---------+ 660
d:
e:                                         EndPheHisIleAspPheTyrThrPro
f:


                        S
                        F       N
                        A       L
                        N       A
                        1       3

         GGTAAAGCATCTAAAGACCTACATGATTATATTCAATCATTAGTTGAGAAAAAATAAATG
     661 ---------+---------+---------+---------+---------+---------+ 720
         CCATTTCGTAGATTTCTGGATGTACTAATATAAGTTAGTAATCAACTCTTTTTTATTTAC

a:       GlyLysAlaSerLysAspLeuHisAspTyrIleGlnSerLeuValGluLysLysEndMet
b:        ValLysHisLeuLysThrTyrMetIleIlePheAsnHisEnd
c:
     661 ---------+---------+---------+---------+---------+---------+ 720
d:                         EndMet                 EndAsnLeuPhePheLeuHis
e:        TyrLeuMetEndLeuGlyValHisAsnTyrGluIleMet
f:
```

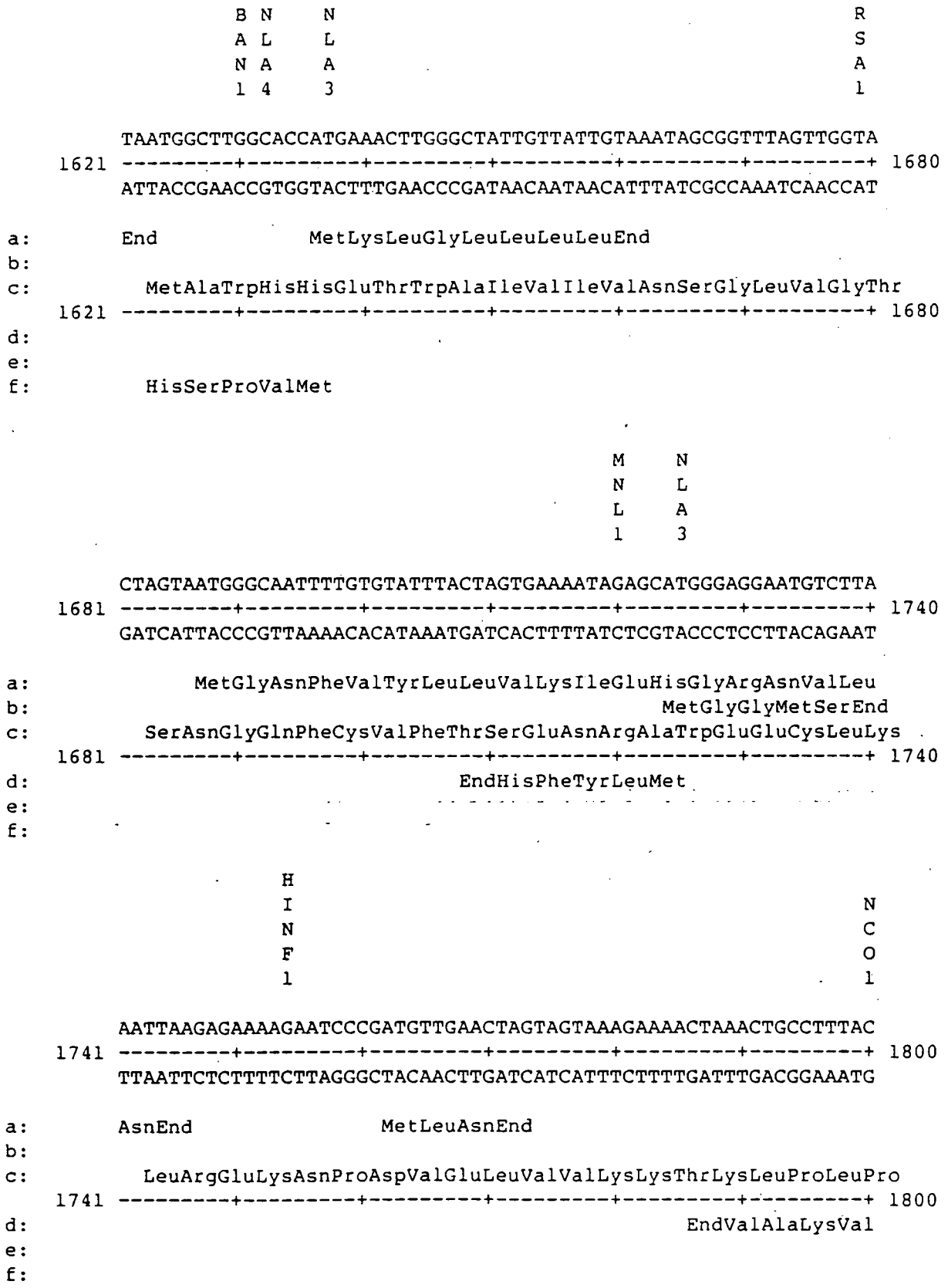

BS SC TR NF 11
RS SC AA 11
B BS AP N1 22
AAAAAGATTATTTTGACTATTGGCTGTCCTGGTTCTGGTAAGAGTACTTGGGCTCGTGAA
721
780
TTTTTCTAATAAAACTGATAACCGACAGGACCAAGACCATTCTCATGAACCCGAGCACTT
a: LysLysIleIleLeuThrIleGlyCysProGlySerGlyLysSerThrTrpAlaArgGlu
b:
c:
721
780
d: PheLeuAsnAsnGlnSerAsnAlaThrArgThrArgThrLeuThrSerProSerThrPhe
e:
f:
H SS
D I AHNNCCS
D N VPCCRRM
E F AAIIFFA
1 1 1211111
TTTATTGCTAAGAATCCCGGGTTTTATAATATCAATCGTGATGACTATCGCCAATCTATT
781
840
AAATAACGATTCTTAGGGCCCAAAATATTATAGTTAGCACTACTGATAGCGGTTAGATAA
a: PheIleAlaLysAsnProGlyPheTyrAsnIleAsnArgAspAspTyrArgGlnSerIle
b: MetThrIleAlaAsnLeuLeu
c:
781
840
d: LysAsnSerLeuIleGlyProLysIleIleAspIleThrIleValIleAlaLeuArgAsn
e:
f: EndArgTrpAspIleIle
H I N P
H H A 1
N L A 3
F N U D 2
M B O 2
R S A 1
A C C 1
ATGGCGCATGAAGAACGCGATGAGTACAAGTATACCAAAAAGAAAGAAGGTATCGTAACT
841
900
TACCGCGTACTTCTTGCGCTACTCATGTTCATATGGTTTTTCTTTCTTCCATAGCATTGA
a: MetAlaHisGluGluArgAspGluTyrLysTyrThrLysLysLysGluGlyIleValThr
b: TrpArgMetLysAsnAlaMetSerThrSerIleProLysArgLysLysValSerEnd
c:
841
900
d: HisArgMet
e:
f: AlaCysSerSerArgSerSerTyrLeuTyrValLeuPhePheSerProIleThrValPro

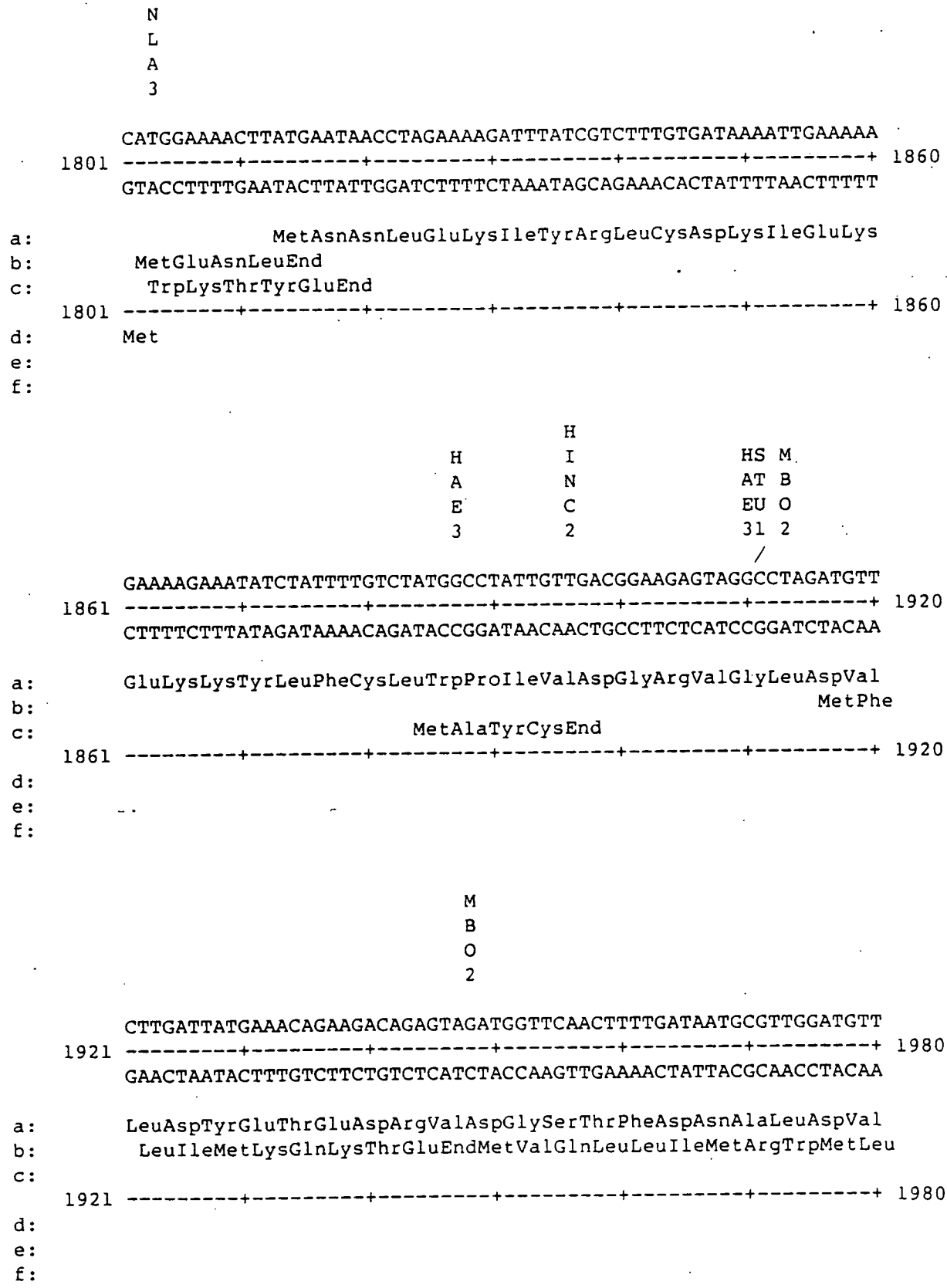
```
                                                  H
                       A            R             I
                       L            S             N
                       U            A             F
                       1            1             1

        GGTATGCAGTTTGATACAGCTAAAAGTATTCTGTACGGTGGCGATTCTGTTAAGGGAGTA
    901 ---------+---------+---------+---------+---------+---------+ 960
        CCATACGTCAAACTATGTCGATTTTCATAAGACATGCCACCGCTAAGACAATTCCCTCAT

a:      GlyMetGlnPheAspThrAlaLysSerIleLeuTyrGlyGlyAspSerValLysGlyVal
b:
c:
    901 ---------+---------+---------+---------+---------+---------+ 960
d:
e:
f:        IleCysAsnSerValAlaLeuLeuIleArgTyrProProSerGluThrLeuProThrIle

                         H
                         I                    N
                         N                    L
                         F                    A
                         1                    3

        ATCATTTCAGATACTAACCTGAATCCTGAACGTCGCCTAGCATGGGAAACTTTTGCCAAA
    961 ---------+---------+---------+---------+---------+---------+ 1020
        TAGTAAAGTCTATGATTGGACTTAGGACTTGCAGCGGATCGTACCCTTTGAAAACGGTTT

a:      IleIleSerAspThrAsnLeuAsnProGluArgArgLeuAlaTrpGluThrPheAlaLys
b:
c:                                                MetGlyAsnPheCysGlnArg
    961 ---------+---------+---------+---------+---------+---------+ 1020
d:                                           EndCysProPheSerLysGlyPhe
e:                    EndGlySerAspGlnValAspGlyLeuMet
f:        MetGluSerValLeuArgPheGlySerArgArgArgAlaHisSerValLysAlaLeuSer


        GAATACGGCTGGAAAGTTGAACATAAAGTGTTTGATGTTCCTTGGACTGAATTGGTTAAA
   1021 ---------+---------+---------+---------+---------+---------+ 1080
        CTTATGCCGACCTTTCAACTTGTATTTCACAAACTACAAGGAACCTGACTTAACCAATTT

a:      GluTyrGlyTrpLysValGluHisLysValPheAspValProTrpThrGluLeuValLys
b:                                          MetPheLeuGlyLeuAsnTrpLeuAsn
c:        IleArgLeuGluSerEnd
   1021 ---------+---------+---------+---------+---------+---------+ 1080
d:      PheValAlaProPheAsnPheMet
e:
f:        TyrProGlnPheThrSerCysLeuThrAsnSerThrGlyGlnValSerAsnThrLeuArg
```


F
N
U
D
2
R
S
A
1
CGTAACTCAAAACGCGGAACTAAAGCAGTACCAATTGATGTTTTACGTTCAATGTATAAA
1081 + 1140
GCATTGAGTTTTGCGCCTTGATTTCGTCATGGTTAACTACAAAATGCAAGTTACATATTT
a: ArgAsnSerLysArgGlyThrLysAlaValProIleAspValLeuArgSerMetTyrLys
b: ValThrGlnAsnAlaGluLeuLysGlnTyrGlnLeuMetPheTyrValGlnCysIleLys
c:
1081 + 1140
d:
e:
f: LeuGluPheArgProValLeuAlaThrGlyIleSerThrLysArgGluIleTyrLeuLeu
H BS
I SC
N TR
F NF
1 11
NS M
LP B
AH O
31 2
AGCATGCGAGAGTATCTCGGTCTTCCAGTATATAATGGGACTCCTGGTAAACCAAAAGCA
1141 + 1200
TCGTACGCTCTCATAGAGCCAGAAGGTCATATATTACCCTGAGGACCATTTGGTTTTCGT
a: SerMetArgGluTyrLeuGlyLeuProValTyrAsnGlyThrProGlyLysProLysAla
b: AlaCysGluSerIleSerValPheGlnTyrIleMetGlyLeuLeuValAsnGlnLysGln
c:
1141 + 1200
d:
e:
f: Met
S
AA
VU
A9
26
R
S
A
1
A
L
U
1
GTTATTTTTGATGTTGATGGTACACTAGCTAAAATGAATGGTCGTGGTCCTTATGACCTT
1201 + 1260
CAATAAAAACTACAACTACCATGTGATCGATTTTACTTACCAGCACCAGGAATACTGGAA
a: ValIlePheAspValAspGlyThrLeuAlaLysMetAsnGlyArgGlyProTyrAspLeu
b: LeuPheLeuMetLeuMetValHisEnd MetValValValLeuMetThrLeu
c:
1201 + 1260
d:
e:
f:

```
                                                                    D
                                                                    D
                                                                    E
                                                                    1

       GAAAAATGCGATACCGATGTTATCAATCCTATGGTTGTTGAACTGTCTAAGATGTATGCT
    1261 ---------+---------+---------+---------+---------+---------+ 1320
       CTTTTTACGCTATGGCTACAATAGTTAGGATACCAACAACTTGACAGATTCTACATACGA

a:     GluLysCysAspThrAspValIleAsnProMetValValGluLeuSerLysMetTyrAla
b:      LysAsnAlaIleProMetLeuSerIleLeuTrpLeuLeuAsnCysLeuArgCysMetLeu
c:          MetArgTyrArgCysTyrGlnSerTyrGlyCysEnd
    1261 ---------+---------+---------+---------+---------+---------+ 1320
d:
e:
f:


       CTTATGGGTTATCAAATCGTAGTCGTTTCAGGTCGTGAAAGTGGAACTAAAGAAGACCCA
    1321 ---------+---------+---------+---------+---------+---------+ 1380
       GAATACCCAATAGTTTAGCATCAGCAAAGTCCAGCACTTTCACCTTGATTTCTTCTGGGT

a:     LeuMetGlyTyrGlnIleValValValSerGlyArgGluSerGlyThrLysGluAspPro
b:      LeuTrpValIleLysSerEnd
c:
    1321 ---------+---------+---------+---------+---------+---------+ 1380
d:
e:                                                        EndLeuLeuGlyLeu
f:


          M                   M
          B                   N
          O                   L
          2                   1

       ACGAAATATTATCGTATGACCCGTAAATGGGTTGAGGACATTGCTGGCGTTCCATTAGTT
    1381 ---------+---------+---------+---------+---------+---------+ 1440
       TGCTTTATAATAGCATACTGGGCATTTACCCAACTCCTGTAACGACCGCAAGGTAATCAA

a:     ThrLysTyrTyrArgMetThrArgLysTrpValGluAspIleAlaGlyValProLeuVal
b:
c:                               MetGlyEnd
    1381 ---------+---------+---------+---------+---------+---------+ 1440
d:                                                              EndAsn
e:      SerIleAsnAspTyrSerGlyTyrIleProGlnProCysGlnGlnArgGluMet
f:            EndArgIleValArgLeuHisThrSerSerMet
```

```
                  H F
                  I NH
                  N UH
                  P DA
                  1 21
                    /
      ATGCAATGTCAGCGCGAACAAGGCGATACCCGTAAAGACGATGTAGTTAAAGAAGAAATT
 1441 ---------+---------+---------+---------+---------+---------+ 1500
      TACGTTACAGTCGCGCTTGTTCCGCTATGGGCATTTCTGCTACATCAATTTCTTCTTTAA

a:    MetGlnCysGlnArgGluGlnGlyAspThrArgLysAspAspValValLysGluGluIle
b:                                                  MetEnd
c:         MetSerAlaArgThrArgArgTyrProEnd
 1441 ---------+---------+---------+---------+---------+---------+ 1500
d:    HisLeuThrLeuAlaPheLeuAlaIleGlyThrPheValIleTyrAsnPhePhePheAsn
e:
f:          EndArgSerCysProSerValArgLeuSerSerThrThrLeuSerSerIleLys


          M                                   A
          B                                   L
          O                                   U
          2                                   1

      TTCTGGAAACACATTGCACCGCATTTTGACGTGAAATTAGCTATTGATGACCGAACTCAA
 1501 ---------+---------+---------+---------+---------+---------+ 1560
      AAGACCTTTGTGTAACGTGGCGTAAAACTGCACTTTAATCGATAACTACTGGCTTGAGTT

a:    PheTrpLysHisIleAlaProHisPheAspValLysLeuAlaIleAspAspArgThrGln
b:                                                  MetThrGluLeuLys
c:
 1501 ---------+---------+---------+---------+---------+---------+ 1560
d:    GluProPheValAsnCysArgMet
e:
f:      GlnPheCysMet

          H         A
          G         H
          A         A
          1         2

      GTAGTTGAAATGTGGCGTCGTATCGGTGTTGAATGCTGGCAAGTCGCTTCGGGAGATTTT
 1561 ---------+---------+---------+---------+---------+---------+ 1620
      CATCAACTTTACACCGCAGCATAGCCACAACTTACGACCGTTCAGCGAAGCCCTCTAAAA

a:    ValValGluMetTrpArgArgIleGlyValGluCysTrpGlnValAlaSerGlyAspPhe
b:     End
c:                                    MetLeuAlaSerArgPheGlyArgPheLeu
 1561 ---------+---------+---------+---------+---------+---------+ 1620
d:
e:
f:                                                              End
```


B N N R
A L L S
N A A A
1 4 3 1
TAATGGCTTGGCACCATGAAACTTGGGCTATTGTTATTGTAAATAGCGGTTTAGTTGGTA
1621 ---------+---------+---------+---------+---------+---------+ 1680
ATTACCGAACCGTGGTACTTTGAACCCGATAACAATAACATTTATCGCCAAATCAACCAT
a: End MetLysLeuGlyLeuLeuLeuLeuEnd
b:
c: MetAlaTrpHisHisGluThrTrpAlaIleValIleValAsnSerGlyLeuValGlyThr
1621 ---------+---------+---------+---------+---------+---------+ 1680
d:
e:
f: HisSerProValMet
M N
N L
L A
1 3
CTAGTAATGGGCAATTTTGTGTATTTACTAGTGAAAATAGAGCATGGGAGGAATGTCTTA
1681 ---------+---------+---------+---------+---------+---------+ 1740
GATCATTACCCGTTAAAACACATAAATGATCACTTTTATCTCGTACCCTCCTTACAGAAT
a: MetGlyAsnPheValTyrLeuLeuValLysIleGluHisGlyArgAsnValLeu
b: MetGlyGlyMetSerEnd
c: SerAsnGlyGlnPheCysValPheThrSerGluAsnArgAlaTrpGluGluCysLeuLys
1681 ---------+---------+---------+---------+---------+---------+ 1740
d: EndHisPheTyrLeuMet
e:
f:
H
I N
N C
F O
1 1
AATTAAGAGAAAAGAATCCCGATGTTGAACTAGTAGTAAAGAAAACTAAACTGCCTTTAC
1741 ---------+---------+---------+---------+---------+---------+ 1800
TTAATTCTCTTTTCTTAGGGCTACAACTTGATCATCATTTCTTTTGATTTGACGGAAATG
a: AsnEnd MetLeuAsnEnd
b:
c: LeuArgGluLysAsnProAspValGluLeuValValLysLysThrLysLeuProLeuPro
1741 ---------+---------+---------+---------+---------+---------+ 1800
d: EndValAlaLysVal
e:
f:


N
L
A
3
CATGGAAAACTTATGAATAACCTAGAAAAGATTTATCGTCTTTGTGATAAAATTGAAAAA
1801
GTACCTTTTGAATACTTATTGGATCTTTTCTAAATAGCAGAAACACTATTTTAACTTTTT
1860
a: MetAsnAsnLeuGluLysIleTyrArgLeuCysAspLysIleGluLys
b: MetGluAsnLeuEnd
c: TrpLysThrTyrGluEnd
1801
1860
d: Met
e:
f:
H
H I HS M
A N AT B
E C EU O
3 2 31 2
GAAAAGAAATATCTATTTTGTCTATGGCCTATTGTTGACGGAAGAGTAGGCCTAGATGTT
1861
CTTTTCTTTATAGATAAAACAGATACCGGATAACAACTGCCTTCTCATCCGGATCTACAA
1920
a: GluLysLysTyrLeuPheCysLeuTrpProIleValAspGlyArgValGlyLeuAspVal
b: MetPhe
c: MetAlaTyrCysEnd
1861
1920
d:
e:
f:
M
B
O
2
CTTGATTATGAAACAGAAGACAGAGTAGATGGTTCAACTTTTGATAATGCGTTGGATGTT
1921
GAACTAATACTTTGTCTTCTGTCTCATCTACCAAGTTGAAAACTATTACGCAACCTACAA
1980
a: LeuAspTyrGluThrGluAspArgValAspGlySerThrPheAspAsnAlaLeuAspVal
b: LeuIleMetLysGlnLysThrGluEndMetValGlnLeuLeuIleMetArgTrpMetLeu
c:
1921
1980
d:
e:
f:

```
                F          M     M
                O          N     B
                K          L     O
                1          1     2

          ATTGATTGGCTTGAAGAAAATTATGTGAGGTAAATATGTTTCCGACTTACTCTAAAATCG
     1981 ---------+---------+---------+---------+---------+---------+ 2040
          TAACTAACCGAACTTCTTTTAATACACTCCATTTATACAAAGGCTGAATGAGATTTTAGC

a:        IleAspTrpLeuGluGluAsnTyrValArgEnd
b:         LeuIleGlyLeuLysLysIleMetEnd
c:                                              MetPheProThrTyrSerLysIleVal
     1981 ---------+---------+---------+---------+---------+---------+ 2040
d:
e:
f:                           EndThrLeuTyrIleHisLysArgSerValArgPheAspTyr

                                                                    F
                                                                    N AP
                                                                    U LV
                                                                    4 UU
                                                                    H 12
                                                                      /
          TAGAAGTAGTGTTTAGCCAAATTATCGCTAATAATATGTTTGAAAAACTTGATAACGCAG
     2041 ---------+---------+---------+---------+---------+---------+ 2100
          ATCTTCATCACAAATCGGTTTAATAGCGATTATTATACAAACTTTTTGAACTATTGCGTC

a:
b:
c:          GluValValPheSerGlnIleIleAlaAsnAsnMetPheGluLysLeuAspAsnAlaAla
     2041 ---------+---------+---------+---------+---------+---------+ 2100
d:
e:
f:          PheTyrHisLysAlaLeuAsnAspSerIleIleHisLysPheLysIleValCysSer

                     H                   H                          BS
          D    A    TBI        N         I        N                 SC
          D    L    ABN        L         N        L                 TR
          E    U    QVF        A         F        A                 NF
          1    1    111        3         1        3                 11
                                                                      /
          CTGAGCTTCGAATCCATGCTCAAGTGACTCATGTATTGAACACTTTGCTTCCAGACCAGG
     2101 ---------+---------+---------+---------+---------+---------+ 2160
          GACTCGAAGCTTAGGTACGAGTTCACTGAGTACATAACTTGTGAAACGAAGGTCTGGTCC

a:                        MetLeuLysEnd    MetTyrEnd
b:
c:          GluLeuArgIleHisAlaGlnValThrHisValLeuAsnThrLeuLeuProAspGlnVal
     2101 ---------+---------+---------+---------+---------+---------+ 2160
d:
e:                                        EndThrAsnPheValLysSerGlySerTrpThr
f:          LeuLysSerAspMetSerLeuHisSerMetTyrGlnValSerGlnLysTrpValLeuHis
```

```
            H                                BS      FH              S
            I                                SC  N   NIH             F
            N                                TR  L   UNH             A
            F                                NF  A   DPA             N
            1                                11  4   211             1
                                             /       /
         TGGATTCTATTGCCATTACACTGTATCCAGGTTCCGCGCATATCATTGTCGTATTTGGTC
    2161 ---------+---------+---------+---------+---------+---------+ 2220
         ACCTAAGATAACGGTAATGTGACATAGGTCCAAGGCGCGTATAGTAACAGCATAAACCAG

a:
b:
c:        AspSerIleAlaIleThrLeuTyrProGlySerAlaHisIleIleValValPheGlyLeu
    2161 ---------+---------+---------+---------+---------+---------+ 2220
d:
e:       SerGluIleAlaMetValSerTyrGlyProGluAlaCysIleMet
f:        IleArgAsnGlyAsnCysGlnIleTrpThrGlyArgMet

                                                        H             E
              D   A                                     I             C
              D   L                                     N             O
              E   U                                     F             R
              1   1                                     1             1

         TTGATGCTGAGCTTGTTATCAAAGGCGATATTCGCTTTGAATCTCAAACTGCGGAATTCA
    2221 ---------+---------+---------+---------+---------+---------+ 2280
         AACTACGACTCGAACAATAGTTTCCGCTATAAGCGAAACTTAGAGTTTGACGCCTTAAGT

a:         MetLeuSerLeuLeuSerLysAlaIlePheAlaLeuAsnLeuLysLeuArgAsnSer
b:
c:        AspAlaGluLeuValIleLysGlyAspIleArgPheGluSerGlnThrAlaGluPheLys
    2221 ---------+---------+---------+---------+---------+---------+ 2280
d:
e:
f:

                                                 A                      M
                                                 L                      N
                                                 U                      L
                                                 1                      1

         AAGCGATTTAATGGTTTACTTTACGGTAGAGCTATAATATCACAACTCTACCAAAACAAA
    2281 ---------+---------+---------+---------+---------+---------+ 2340
         TTCGCTAAATTACCAAATGAAATGCCATCTCGATATTATAGTGTTGAGATGGTTTTGTTT

a:       LysArgPheAsnGlyLeuLeuTyrGlyArgAlaIleIleSerGlnLeuTyrGlnAsnLys
b:                 MetValTyrPheThrValGluLeuEnd
c:        AlaIleEnd                                                  Met
    2281 ---------+---------+---------+---------+---------+---------+ 2340
d:                                                 EndLeuGluValLeuValPhe
e:
f:
```

```
                                              E
                    D                         C
                    D                         O
                    E                         R
                    1                         1

     TGAGGAAAACAAAATGTTGCTAAGTGAAAAACCGATTACTGTTAAAGGAATTC
2341 ---------+---------+---------+---------+---------+--- 2393
     ACTCCTTTTGTTTTACAACGATTCACTTTTTGGCTAATGACAATTTCCTTAAG

a:   End
b:             MetLeuLeuSerGluLysProIleThrValLysGlyIle???
c:     ArgLysThrLysCysCysEnd
2341 ---------+---------+---------+---------+---------+--- 2393
d:   SerSerPheLeuIleAsnSerLeuSerPheGlyIleValThrLeuProIle???
e:                   EndThrPhePheArgAsnSerAsnPheSerAsn???
f:                                         EndLeuPheGlu
```