The EMBO Journal vol.4 no.10 pp.2695-2703, 1985

# T4 polynucleotide kinase; cloning of the gene (*pse*T) and amplification of its product

Carol A.Midgley and Noreen E.Murray

Department of Molecular Biology, King's Buildings, Edinburgh University, Mayfield Road, Edinburgh EH9 3JR, UK

Communicated by N.E.Murray

**The T4 gene (*pse*T) for polynucleotide kinase (pnk) has been cloned in λ. Induction of a λ$E^-W^-S^-$cI857 prophage in which the *pse*T gene can be transcribed from the late λ promoter, $p_{R'}$, leads to >100-fold amplification of pnk activity; pnk comprises ~7% of the total soluble cell protein. The purified enzyme, as expected, is both a 5'-kinase and a 3'-phosphatase. The amino acid sequence deduced from an open reading frame identified as the *pse*T gene contains a sequence which corresponds particularly well with that part of the adenine nucleotide binding site of adenylate kinase shown to form a flexible loop. A deletion mutant that lacks 5'-kinase activity, and possibly also 3'-phosphatase activity, has lost two amino acids from within the proposed loop structure. A second region of the pnk sequence shares homology with phosphoglycerate kinase, yeast inorganic pyrophosphatase and histone 2b from various organisms.**

*Key words:* enzyme amplification/gene cloning/nucleotide binding site

## Introduction

Polynucleotide kinase (pnk), the product of the T4 early gene *pse*T (Sirotkin *et al.*, 1978), catalyses the transfer of the γ-phosphate of ATP to a 5'-hydroxyl terminus of DNA or RNA (Richardson, 1965; Novogrodsky and Hurwitz, 1966) or oligonucleotides, the minimal substrate sequence being a nucleotide 3'-phosphate. The enzyme also has a 3'-phosphatase activity which strongly prefers DNA as a substrate (Cameron and Uhlenbeck, 1977; Sirotkin *et al.*, 1978).

The *pse*T gene is non-essential since phage in which it is deleted grow well in most *Escherichia coli* strains. However, a clinical strain non-permissive for *pse*T$^-$ phage has facilitated screening for mutants, and the subsequent genetic analysis of the *pse*T locus (Depew and Cozzarelli, 1974). Most mutations in *pse*T inactivate both the 5'-kinase and the 3'-phosphatase although the point mutation *pse*T1 is known to inactivate only the 3'-phosphatase (Sirotkin *et al.*, 1978).

The *in vivo* role of the two enzymic activities has not been determined. David *et al.* (1982a,1982b) have suggested that both functions may be involved in RNA processing since their action may produce appropriate substrates for T4 RNA ligase, the product of gene 63 which maps close to *pse*T (Snopek *et al.*, 1977). Additional interest in RNA processing in prokaryotes has been stimulated recently by the discovery of an intron in the T4 thymidilate synthetase (*td*) gene (Chu *et al.*, 1984). Other evidence, however, suggests that pnk is involved in DNA synthesis since *pse*T mutations reduce the rate of DNA replication and also the length of the DNA molecule synthesised in a permissive host (Runnels *et al.*, 1982; Sirotkin *et al.*, 1978).

In addition to the interest in the role of pnk *in vivo*, there is widespread use of the enzyme in experimental molecular biology. A system of producing the enzyme in the absence of contaminating T4 nucleases, and of increasing the level of expression would improve the purification procedure.

The locus of the *pse*T gene is close to gene 63, and between the *td* and DNA ligase genes (Depew and Cozzarelli, 1974). Mileham *et al.* (1980) used λ derivatives carrying the T4 *td* and DNA ligase genes as probes to identify an 11.5-kb *Hind*III fragment which links these two markers and includes the *pse*T region. Although this large fragment could not be cloned, almost all of the component *Eco*RI fragments were isolated in λ vectors. The approximate location of *pse*T is defined by a T4 deletion mutant *pse*TΔ1 (see Figure 1) which is known to delete all *pse*T markers. This deletion removes all or most of the contiguous 2.27-kb, 0.1-kb and 1.1-kb *Eco*RI fragments (Mileham *et al.*, 1980). An analysis of heteroduplex molecules indicates that *pse*TΔ1 is 3.5 kb in length (Kutter *et al.*, 1984), so it is unlikely to extend very far outside these three *Eco*RI fragments.

The mol. wt. of pnk as determined by gel filtration is 140 kd, whereas the polypeptide separated by SDS-gel electrophoresis has a mol. wt. of ~33 kd (Panet *et al.*, 1973). Lillehaug (1977) suggested that the active pnk consists of four identical subunits since the only N-terminal amino acid is phenylalanine.

The entire coding sequence for a polypeptide of mol. wt. 33 kd could be contained in either of the two larger fragments missing from T4 *pse*TΔ1. Alternatively, the coding sequence could span more than one fragment, or extend outside the region defined by *pse*TΔ1. This paper defines the coding sequence for the *pse*T gene within the 2.27-kb *Eco*RI fragment, and shows how λ clones including this fragment may be used to amplify the yield of pnk.

## Results

### *T4 EcoRI fragment 21 encodes a polypeptide of mol. wt. 33 kd*

Mileham *et al.* (1980) cloned *Eco*RI fragments from the frd-DNA ligase region of phage T4 in the λ insertion vector NM607. Their clones, none of which contained more than one *Eco*RI fragment, included fragments 21 (2.27 kb), 33 (1.1 kb), and 46 (0.1 kb), which are all missing from T4 *pse*TΔ1 (see Figure 1). Recombinant derivatives of vector NM607 are readily selected, but these integration-deficient, c$I^-$ phages are unable to lysogenise, and while the cloned DNA fragment may be transcribed leftwards from $p_{RE}$ (see Figure 3), rightwards transcription from a λ promoter has not been documented. Alternative λ vectors allow propagation in the prophage state and offer greater potential for controlled transcription irrespective of the orientation of the insert. Fragment 21 was, therefore, transferred from the original λ clone (NM1210) to the central region of an integration-proficient λ vector, NM459 (see NM1104 in Figure 3). Polypeptides were analysed following infection of u.v.-irradiated bacteria (M159) with either λ or T4 phage. The λ phage NM1104 and its *Wam Eam Sam* derivative NM1108 (see Figure 2), but not the vector, expressed a polypeptide of mol. wt. 33 kd. This

polypeptide migrated in a position corresponding to the *pse*T polypeptide, which is identified by its absence from cells infected with the T4 *pse*T$^-$ mutant *pse*T2 (Figure 2).

The orientation of fragment 21 in NM1104 was readily deduced from the position of the asymmetrically located *Sma*I site (Figure 3). In this phage, and its derivatives, leftwards transcription of the T4 DNA is mediated early from the powerful λ promoter $p_L$ and rightwards transcription is dependent on subsequent activation of the late λ promoter $p_{R'}$ (see Figure 3). The direction of transcription of the coding sequence for the 33-kd polypeptide was deduced from the timing of expression following infection of u.v.-irradiated cells. In the case of NM1104 or NM1108, little expression of the 33-kd polypeptide was detected when labelling was terminated 13 min after infection. Much more was detected when proteins were labelled 20 min after infection (see tracks d and e in Figure 2). This implies that transcription of the coding strand needs, or is enhanced by, activation of the late promoter $p_{R'}$. A second recombinant, CM1, which has fragment 21 in the opposite orientation (see Figure 3), expressed the 33-kd polypeptide more effectively during the first 13 min following infection, as expected for transcription from the early λ promoter $p_L$ (data not shown). The timing of expression identifies the orientation of the sequence encoding the 33-kd polypeptide relative to the λ promoters. Therefore, by extrapolation with respect to markers on the T4 map the coding sequence is transcribed anti-clockwise with respect to the T4 genome (i.e., from left to right as shown in Figure 1), as expected for a gene expressed early in the T4 life cycle.

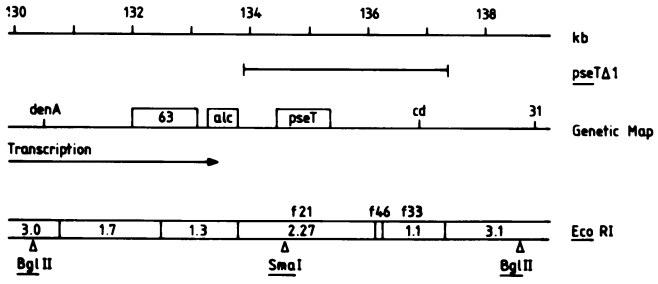


**Fig. 1.** Organisation of the *pse*T region. The figure shows: physical distance in kilobase pairs from the rIIA - rIIB join = 0 (Wood and Revel, 1976); the approximate map position of T4 deletion *pse*TΔ1 (Mileham *et al.*, 1980); a genetic map of T4 showing the gene order, rectangles represent genes whose lengths have been determined by DNA sequence; the direction of transcription; the numbers allotted to the relevant *Eco*RI fragments (Kutter *et al.*, 1984); restriction sites for *Eco*RI, *Bgl*II and *Sma*I.

Some idea of the location of the sequence encoding the 33-kd polypeptide was gained from λ CM6, a recombinant from which the DNA upstream of the *Sma*I site was deleted (see Figure 3). This phage failed to express a 33-kd polypeptide (data not shown), suggesting that the *Sma*I site is within the sequence coding for this polypeptide.

Crude extracts from non-irradiated cells infected with phage were assayed for 5′-kinase activity (see Materials and methods).

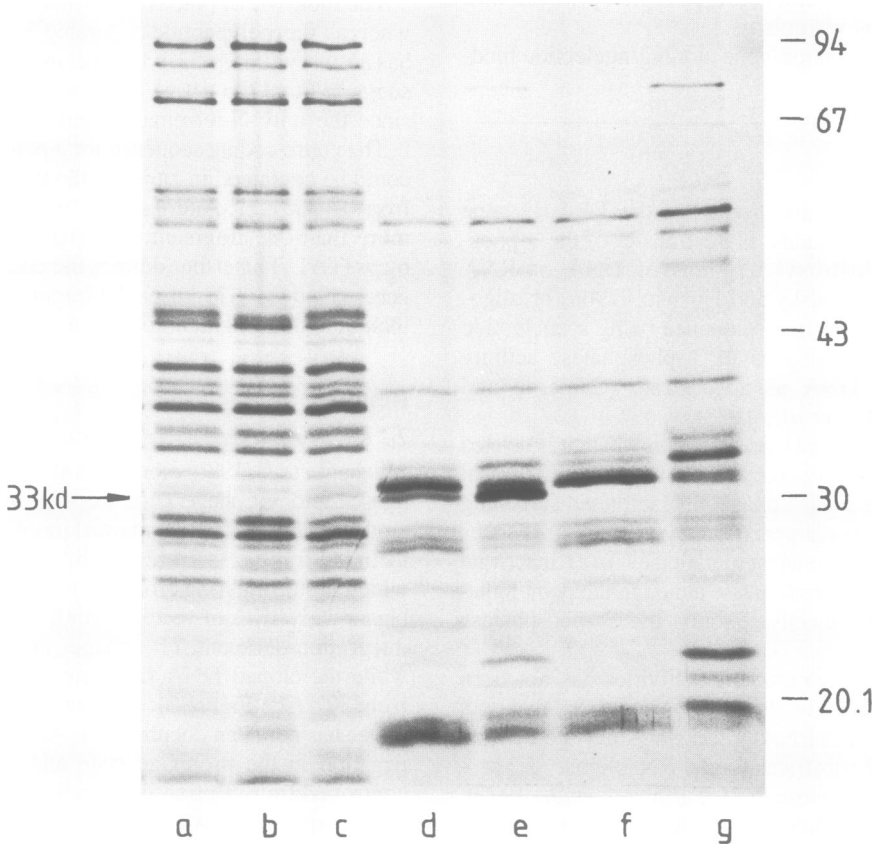


**Fig. 2.** Autoradiographic analysis of $^{35}$S-labelled polypeptides. Polypeptides were separated by electrophoresis through a 15% linear SDS-polyacrylamide gel (see Materials and methods). Infections with T4 were labelled from 3 to 7′ **(tracks a, b** and **c)**, those with λ derivatives from 3 to 13′ **(tracks d** and **f)**, or 20 to 30′ **(tracks e** and **g)**. The positions of standard marker proteins are indicated: phosphorylase b (94 kd); bovine serum albumin (67 kd); ovalbumin (43 kd); carbonic anhydrase (30 kd); and soybean trypsin inhibitor (20 kd). The tracks are: **a**, T4D (wild-type); **b**, T4 *pse*T2 a *pse*T$^-$ strain; **c**, T4 *am*N122; **d**, λNM1108 labelled 3−13′; **e**, λNM1108 labelled 20−30′; **f**, λNM1108 deleted for *Eco*RI fragment 21 labelled 3−13′; **g**, λNM1108 deleted for *Eco*RI fragment 21 labelled 20−30′. The position of the mol. wt. 33 kd polypeptide is marked by an arrow.

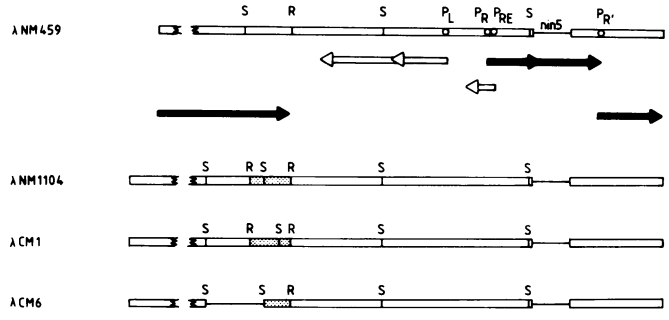


**Fig. 3.** Transcription patterns of vector NM459. The vector has a single *Eco*RI site (R) to the left of the attachment site. The genome circularises on infection and fragments inserted at the *Eco*RI site may be transcribed early from the leftwards λ promoter $p_L$ and late from the rightwards λ promoter $p_{R'}$. The transcript from $p_{R'}$ not only requires activation by pQ, a product of early rightwards transcription, but must traverse >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 fragment in NM1104 and CM1 were determined from *Sma*I digests (S indicates *Sma*I sites). λCM6 is like λCM1 but is deleted for the DNA between the two leftmost *Sma*I sites; this phage no longer expresses the 33-K polypeptide (data not shown). The open arrows show major λ leftwards transcripts, the solid arrows the major rightwards transcripts, and T4 DNA is shaded.

A correlation was observed between 5′-kinase activity and the presence of a *pse*T polypeptide in T4 strains, but no activity was detected in extracts from suppressor-free cells (ED8689) infected with NM1108, a phage that on infection of u.v.-irradiated cells produced large amounts of the 33-kd polypeptide (see Figure 2). Therefore, if the gene in fragment 21 encoding the 33-kd polypeptide is *pse*T, the resulting protein is defective in pnk activity. A defective protein could result from a mutation, or because part of the coding sequence is outside fragment 21. A further possibility is that some essential processing of the mRNA or of the polypeptide fails to occur for the λ clone. We favoured one of these explanations rather than the possibility that the complete *pse*T gene is located outside fragment 21, since there is no obvious alternative 33-kd polypeptide encoded by T4$^+$, and T4*pse*TΔ1 phage infections are reported to lack only one 33-kd polypeptide (E.Kutter, personal communication).

*The DNA sequence of EcoRI fragments 21 and 46*

*Eco*RI fragments 21 and 46, derived from λNM1210 and λNM1207, respectively (Table I), were transferred to M13 vectors. The entire fragment 46 and the extreme ends of fragment 21 were sequenced by the dideoxy chain termination method. DNA sequence for fragment 21 was obtained mainly from subclones generated in M13 by sonication of the purified fragment. The sequences of random fragments were compiled and analysed using the computer programs of Staden (1982).

The orientation of fragment 21 in relation to the T4 physical and genetic map (Figure 1) is indicated by the unique *Sma*I site. The orientation of fragment 46 was determined by the DNA sequence of an *Alu*I fragment [derived from an *Alu*I digest of the 8.2-kb *Bgl*II fragment (Figure 1)] which hybridised to both fragment 46 and fragment 21. Sections of DNA sequence not initially obtained on both strands from the sequence of random fragments were completed by selection of the appropriate clones from the library of random fragments in M13 using previously sequenced clones as single-stranded hybridisation probes (Hu and Messing, 1982). One final section was sequenced by priming M13 clones carrying the entire fragment 21 with a specific synthetic oligonucleotide.

**Table I.** Phage strains

| Strains | Relevant features and use | Source or reference |
|---|---|---|
| (a) λ vectors and their derivatives | | |
| NM607 | Immunity insertion vector for *Eco*RI fragments | Murray *et al.* (1977) |
| NM1207 | NM607 including T4 fragment 46[a] | Mileham *et al.* (1980) |
| NM1210 | NM607 including T4 fragment 21[a] | Mileham *et al.* (1980) |
| NM459 | (*sr*Iλ1-2)Δ *c*I857 *nin*5 | See Figure 3 |
| NM1104 | NM459 including T4 fragment 21 | See Figure 3 |
| CM1 | NM459 with fragment 21 in opposite orientation | See Figure 3 |
| CM6 | Deletion derivative of NM1104 | See Figure 3 |
| NM1108 | *Wam*403 *Eam*1100 *Sam*100 derivative of NM1104 | |
| NM1149 | Immunity insertion vector for *Eco*RI or *Hind*III fragments | Murray (1983) |
| CM8 | NM1149 including fragment 21[b] | See Figure 5 |
| NM1070 | *Wam*403 *Eam*1100 *lacZ* *c*I857 *nin*5 *Sam*100 | See Figure 5 |
| CM21 | NM1070 including fragment 21[b] in place of *lacZ* | See Figure 5 |
| (b) T4 phages | | |
| T4D | T4$^+$ | H.R.Revel (Personal communication) |
| *pse*T2 | Defective in polynucleotide kinase polypeptides | Sirotkin *et al.* (1978) |
| *am*N122 | g42$^-$; *pse*T$^+$ | Hughes and Brown (1973) |
| *alc*7 | *alc*7, *am*51(g56), NB5060(*den*B-*r*II); *am*87(g42); 100% cytosine DNA | Wilson *et al.* (1977) |

[a]See Figure 1.
[b]Alternative clone of fragment 21.

The DNA sequence (Figure 4) revealed only one large open reading frame (ORF) of 900 bases which was entirely within fragment 21. Several smaller ORFs lie within the sequenced region, but no large ORF extends over fragments 21 and 46 and into fragment 33, since there are stop codons in all reading frames in fragment 46. The direction of transcription required for the 900-base ORF corresponds to that required for expression of the 33-kd polypeptide in the λ phages NM1108 and NM1104. The proposed ATG initiation codon for the ORF is situated 75 bases upstream of the *Sma*I site. This is consistent with the loss of expression of the 33-kd polypeptide in the λ recombinant (CM6) deleted for the smaller of the *Sma*I-*Eco*RI fragments derived from fragment 21 (Figure 3). However, although phenylalanine has been reported as the N-terminal amino acid of pnk purified from T4-infected cells (Lillehaug, 1977) the first phenylalanine encoded by the ORF is at residue 20. Therefore, if this ORF does indeed encode the pnk polypeptide, the lack of kinase activity may reflect a failure of post-translational processing in the absence of some T4-encoded function. An equally plausible explanation is a mutation within the cloned ORF. For this reason, additional clones were made in which fragment 21 came from a different DNA preparation.

*Some clones including fragment 21 encode an active polynucleotide kinase*

The 8.2-kb *Bgl*II fragment (Figure 1) was purified from T4 *alc*7 DNA, digested with *Eco*RI and the resulting fragments were cloned in the λ vector NM1149 (this vector differs from NM607 only in that it may be used for either *Eco*RI or *Hind*III fragments). λ recombinants containing fragment 21 were identified by hybridi-


**Fig. 4.** DNA sequence of the contiguous *Eco*RI fragments 21 and 46 and the protein sequence of the ORF encoding polynucleotide kinase. The DNA sequence includes the 6 bp, GTC CTG (underlined) deleted in the sequence encoding the inactive mol. wt. 33 kd polypeptide. Some possible −10 and −35 promoter regions are indicated, but the −10 region immediately upstream of the initiation codon lacks a −35 region. The initiation codon is not preceded by a perfect Shine-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.

sation to purified fragment 21. Crude extracts from cells infected with these recombinants, in contrast to those infected with NM1108, were found to have considerable 5′-kinase activity.

Fragment 21 was purified from one of the new λ-T4 recombinants (λCM8). *Alu*I fragments were subcloned in M13 and the ORF was re-sequenced. The sequence differed from that obtained for the kinase-deficient clone only by the addition of six bases, GTC CTG (bases 746−751 in Figure 4), upstream of the first phenylalanine. The sequence of the large ORF given in Figure 4 includes the extra six bases and is identical to that obtained by K. Rand and M. Gait (personal communication) for T4 DNA cloned directly in M13. The mol. wt. of the polypeptide encoded by the 906-bp ORF is calculated to be 34 kd.

*Purification of the 33-kd polypeptide and characterisation of enzyme activity*

Active pnk was purified (Materials and methods) from cells infected with λCM8. The purified protein was incubated in the presence of [$\gamma$-$^{32}$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 T1 digest of yeast tRNA. Incorporation of [$^{32}$P]phosphate into both species demonstrated kinase activity, while a subsequent venom phosphodiesterase digest of the labelled oligodeoxyribonucleotides showed that label had been incorporated into all four deoxyribonucleotides. Digestion of 3′-[$^{32}$P]-

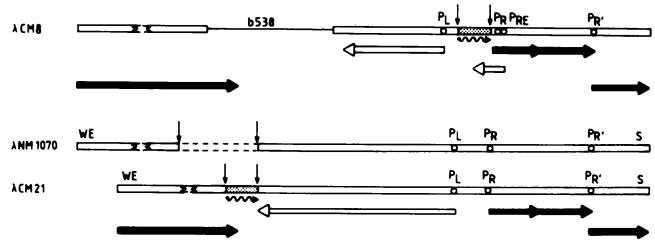


**Fig. 5.** The origin and transcription of λCM21. The structure of λCM8, the first recombinant encoding an active polynucleotide kinase, is shown. Fragment 21 is inserted in the orientation in which *pseT* cannot be transcribed from $p_{RE}$. Transcription of the *pseT* gene may depend on a promoter sequence within the insert, although transcription from a λ promoter has not been ruled out. In λCM21, T4 fragment 21 replaces the central *lacZ* fragment of vector NM1070. 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 the *Eco*RI 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 both phages was determined from *Sma*I digests). Large open arrows show major leftwards transcripts, solid arrows the rightwards transcripts.

mononucleotides with the purified enzyme indicated the predicted 3′-phosphatase activity. The 33-kd polypeptide purified from cells infected with NM1104 lacked 3′-phosphatase activity; assays of crude extracts of infected cells have not been done.

*Amplification of T4 polynucleotide kinase following induction of a prophage*

The enzymic activities of pnk purified from cells infected with λCM8 were indistinguishable from those of the T4-encoded enzyme. However, λCM8 does not provide a convenient source for routine preparations of enzyme since phage lysates are required to infect bacteria and the infected cells must be harvested before they lyse. One means of amplifying pnk would be to construct a series of fusions placing the *pseT* gene downstream of a strong controllable promoter in a multicopy plasmid. An alternative approach, as used for T4 DNA ligase, is to position fragment 21 in the centre of an integration-proficient, temperature-inducible λ vector that is defective in both lysis and packaging (Murray *et al.*, 1979). The λ vector provides the simpler method and fragment 21 was transferred from λCM8 to λ*Wam Eam lacZ att*$^+$ *int*$^+$ *c*I857 *Sam* (NM1070) in place of the *lacZ* fragment (Figure 5). Lysates were made from those phages of the correct genotype which hybridised to purified fragment 21. These phages gave minute plaques and, although it was not possible to make really high titre lysates, assays of extracts of infected cells showed that polynucleotide kinase was produced. One of these phages, λCM21, was chosen and lysogens of the suppressor-free host, ED8689, were isolated. On induction, such a lysogen was found to yield high levels of pnk activity.

A lysogen will probably have one copy of λCM21 per host chromosome, but following induction the λ genome will replicate to give 100 or more copies per bacterium. Even if the cloned *pseT* gene lacks a functional promoter sequence, transcription of this amplified gene will result following activation of the late λ promoter $p_{R'}$ (see Figure 5).

Extracts of the induced lysogen, ED8689 (λCM21), and of ED8689 infected with T4$^+$, were diluted serially and assayed for 5′-kinase activity by measuring conversion of the dinucleoside monophosphate CpU to the dinucleotide pCpU (Figure 6). In assays on crude cell extracts a nuclease activity degrades pCpU to pC (K.Murray, personal communication). Consequently a radiolabelled spot migrating with the mobility of 5′-dCMP is seen on paper ionophoresis instead of the expected product 5′-[$^{32}$P]-pCpU (Figure 6). This spot is gradually lost as the T4$^+$ extract is diluted. In contrast, when reactions are carried out with serial dilutions of extracts of the induced lysogen, the pC spot is replaced by one migrating in the position of pCpU. This spot is only lost after at least a 2000-fold dilution of the extract. Amplification of pnk activity appears to be sufficiently great that kinase activity remains after the nuclease activity has been lost by dilution. In concentrated crude extracts (particularly tracks j−1 in Figure 6) there is sufficient kinase activity to use all the [$\gamma$-$^{32}$P]ATP, and in the most concentrated extract (track j, Figure 6) limiting donor in the presence of the nuclease prevents the accumulation of any pCpU.

Although the assay used is not directly quantitative, comparison of the activities in extracts of T4$^+$-infected cells and the induced lysogen, even when normalised with respect to total protein concentration, shows that the lysogen yields at least a 100-fold more activity than does infection by T4$^+$. The level of pnk produced on induction of the prophage has not been accurately quantified, but a rough estimate based upon densitometer scans of acrylamide gels indicates that it constitutes ~7% of the total protein in crude extracts of induced lysogens (Figure 7).

## Discussion

Identification of the coding sequence for polynucleotide kinase was impeded by the fact that fragment 21 of the T4 genome, as originally cloned in a λ vector by Mileham *et al.* (1980), carried a defective *pseT* gene. This defective clone misled us into overestimating the difficulties involved in cloning an active *pseT* gene. In reality the active gene was readily cloned in a λ immunity insertion vector, but only with difficulty in the central region of a λ vector where expression will be enhanced by transcription from λ promoters (see Figure 5).

Amplification of polynucleotide kinase, like amplification of some other enzymes that affect DNA metabolism, e.g., DNA polymerase I (Kelley *et al.*, 1977) and DNA ligase (Wilson and Murray, 1979), appears to be deleterious to the host. In this context it is probably significant that, regardless of the λ vector used, fragment 21 has been cloned in only one orientation; that avoiding early expression. For those phages where we anticipate effective transcription of *pseT* from $p_{R'}$ (e.g., CM21 in Figure 5) plaques are minute, and lysates difficult to prepare. A high titre for DNA preparation was achieved only following induction of a lysogen and concentration of cells before lysis. Propagation of CM21 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, particularly if there is no strong promoter in fragment 21 upstream of *pseT*. After seven successive subcultures in which the lysogen was diluted 1000-fold and grown to saturation, >90% of bacteria remained lysogenic and those colonies tested gave high levels of pnk activity.

We have no explanation for the occurrence of a 6-bp deletion in the *pseT* gene of our original clone. There is no obvious feature in the DNA sequence which might stimulate deletion during phage propagation. 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. For example, it was never possible to lysogenise *E. coli* with an integration-proficient phage carrying the defective *pseT* gene and it was

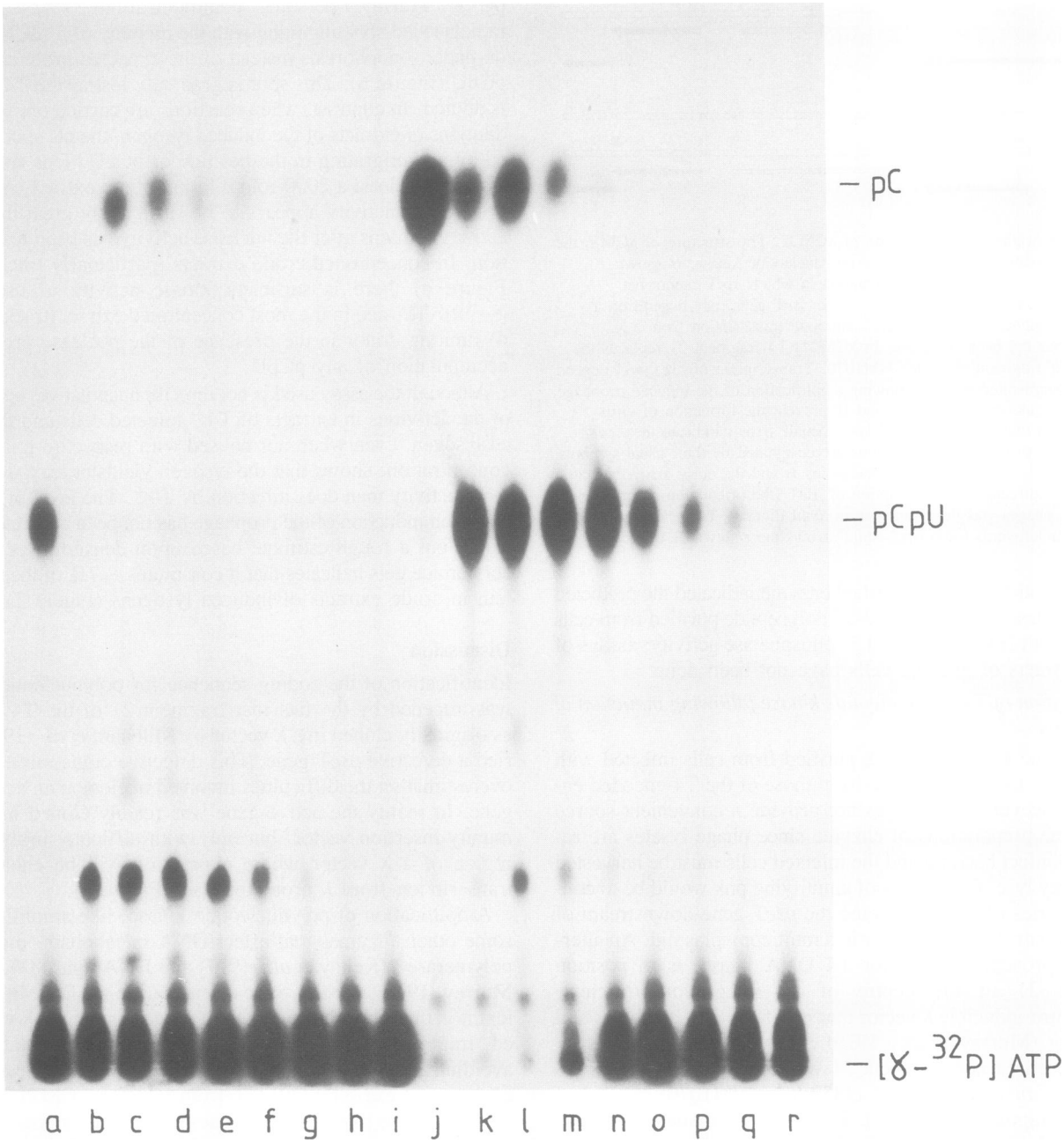


**Fig. 6.** Assay for 5′-kinase activity in crude cell extracts. The assay involves the conversion of the substrate CpU to the dinucleotide pCpU by the incorporation of phosphate from [$\gamma$-$^{32}$P]ATP (see Materials and methods and Results). The positions of pC and pCpU species are indicated. **Track a**, a control using purified polynucleotide kinase enzyme from cells infected with T4amN122. Other assays: **track b**, cells infected with T4 *pse*T2 (*pse*T$^-$); **tracks c – i**, 2-fold serial dilutions of infections with T4D (wild-type); **tracks j – r**, 3-fold serial dilutions of the induced lysogen ED8689 (CM21). The undiluted crude cell extracts used in **tracks c** and **j** were adjusted to have the same total protein concentration as determined by Lowry-Folin assays (Lowry *et al.*, 1951). A nuclease in crude extracts (see text) degrades pCpU to pC; this nuclease activity is detected when purified pnk is added to crude cell extracts.

difficult to grow or maintain a culture of *E. coli* carrying a plasmid containing the defective fragment 21.

A search for homologies between the amino acid sequence encoded by the *pse*T gene and the NBRF protein sequence database revealed a striking homology with the N-terminal region of adenylate kinase, an enzyme that catalyses the phosphorylation of AMP by ATP. Comparison with other adenine nucleotide-binding proteins (Walker *et al.*, 1982) substantiates the presence in the active pnk polypeptide of a conserved sequence characteristic of an adenine nucleotide binding site (Table II). In adenylate kinase, residues 16 – 23 (Table IIA) form a flexible loop structure of sequence Gly-X-Gly-X-Gly. Pai *et al.* (1977) using crystallographic techniques showed that this loop forms part of the AMP-specific adenine nucleotide binding site, and that the conformation of the loop changes dramatically when AMP is bound. The lysine following the loop (Lys21) may interact with the $\alpha$-phosphate of AMP and appears to be a conserved feature of adenine nucleotide binding sites. A similar sequence is seen in the ATP synthetases and in active pnk (Table IIA) which suggests that residues 10 – 17 of pnk form a loop structure analogous to that described for adenylate kinase. Deletion of Pro11 and Gly12 in the inactive pnk polypeptide would constrict the postulated loop and perhaps prevent binding of adenine nucleotides or block a conformational change initiated by the binding of an

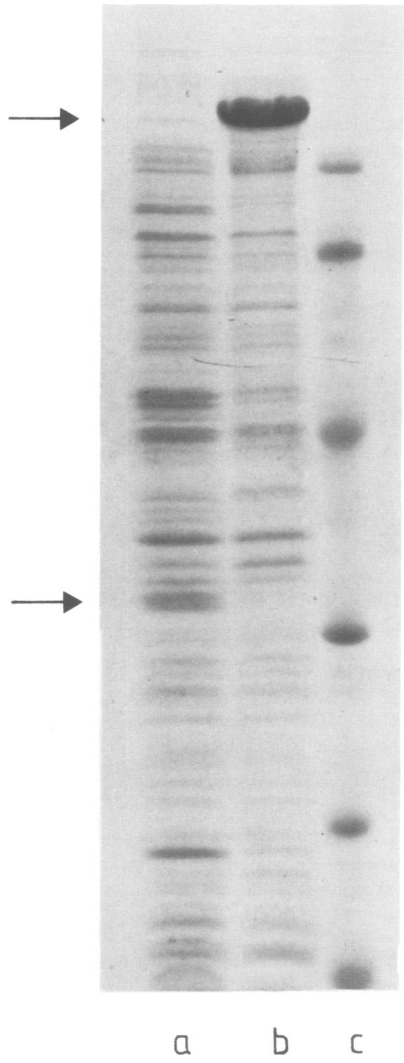


**Fig. 7.** Analysis of polypeptides after induction of a λ prophage including the *pse*T gene. The polypeptides were separated on an SDS-polyacrylamide gel and stained with Coomassie blue. **Track a**, crude extracts from induced ED8689 (λCM21) **b**, crude extracts from induced ED8689 (λNM1070) **c**, standard markers, see legend to Figure 2. The top arrow indicates β-galactosidase expressed by the vector NM1070 **(track b)**, the lower arrow pnk expressed by λCM21 **(track a)**. In this gel pnk migrates as a broad (doublet) band due to insufficient reducing agent. On a subsequent gel in the presence of a higher concentration of mercaptoethanol a well-defined single band was obtained.

adenine nucleotide. We have no experimental evidence to support either of these suggestions and cannot say whether a change in the postulated loop would be directly responsible for the loss of both 5′-kinase and 3′-phosphatase activities. The overall conformation of the defective pnk seems unlikely to be markedly different from that of the active enzyme, since the two proteins purify in the same way and there is no significant difference in their predicted secondary structures (L.Sawyer, personal communication).

We have not detected homology between pnk and a second region of sequence associated with AMP binding in adenylate kinase (Walker *et al.*, 1982). However, residues 127–131 (Table IIB) have homology with sequences in human and horse phosphoglycerate kinase and with yeast inorganic pyrophosphatase. This same sequence is also conserved in histone 2b from various species. The possible relevance of this second conserved sequence to nucleotide binding remains to be elucidated.

The DNA sequence of the *pse*T gene (Figure 6) is not consistent with the reported N terminus for pnk (Lillehaug, 1977). We initially postulated that post-translational processing may be necessary, but since the first phenylalanine is beyond the proposed adenine nucleotide binding sequence, processing by addition of phenylalanine rather than the removal of amino acids would be required. If processing is a requirement for enzymic activity, it is achieved in the absence of T4 functions other than those encoded by fragment 21. The amino acid sequence of the wild-type polypeptide will show whether processing does occur and, if so, it will be of interest to see whether processing is affected by the change in amino acid sequence associated with the deletion mutant.

The level of enzyme activity produced by our prophage system is >100 times that in extracts of T4$^+$-infected cells, and for the first time pnk can be prepared without making phage lysates. Of the four T4-encoded enzymes that are commonly used in the laboratory, DNA ligase (Murray *et al.*, 1979), RNA ligase (Rand and Gait, 1984) and polynucleotide kinase can now be prepared without recourse to T4-infected cells. The considerable amplification of pnk is of obvious relevance to its production as a basic reagent, and to detailed studies of both its mechanism of action and its role *in vivo*.

## Materials and methods

*Strains, vectors and media*

Bacterial strains are listed in Table III; λ and T4 phages, including λ vectors, in Table I. M13mp10, mp11 and mp18 (Messing, 1983) were used as vectors for DNA sequencing. A derivative of plasmid pBR325 (Bolivar, 1978) including

**Table II.** Homologies with pnk amino acid sequence

| | 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 |
| | 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. | Histone 2b (e.g. sea urchin) | 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 (human/horse) | 338–363 | P V G V F E W E A F A R [G T K A] [L] M D E V V K A T S |
| | Inorganic phosphatase (bakers yeast) | 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 |

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) (Walker *et al.*, 1982) is indicated by boxes.
B shows a second region of homology of undetermined significance (see Discussion).

**Table III.** Bacterial strains

| *E. coli* K12 | Relevant features | Use | Source and/or reference |
|---|---|---|---|
| ED8689 | *sup°* *hsdR* | Non-permissive host for amber mutants | Wilson *et al.* (1977) |
| CR63 | *supD* | Permissive host for T4 amber mutants | Appleyard *et al.* (1956) |
| NM538 | *supF hsdR* | Permissive host for λ amber mutants | Frischauf *et al.* (1983) |
| NM514 | *hfl hsdR* | Selection of $cI^-$ recombinants | Murray (1983) |
| C600 | *supE* | Amplification of λ*Wam Eam Sam* phages | Appleyard (1954) |
| M159 | *sup°* *uvrA* | Host for analysis of polypeptides | Jaskunas *et al.* (1975) |
| NM522 | (*lac-pro*)Δ*hsdMS*Δ F′*lacZ*M15*lacI*q | Host for M13 phages | Gough and Murray (1983) |

T4 *Eco*RI fragment 46 (see Figure 1) was used as a hybridisation probe. Media were made and used as described by Murray *et al.* (1977).

*Enzymes and chemicals*

DNA polymerase (Klenow fragment) was purchased from Boehringer, Mannheim, T4 DNA ligase from New England Biolabs Inc. Beverly, MA, DNA polymerase I from NBL Enzymes Ltd., Cramlington, UK, restriction enzymes from New England Biolabs Inc. Beverly, MA, or Bethesda Research Laboratories (UK) Ltd. Cambridge, UK, RNase A and T1, calf intestinal phosphatase, lysozyme and micrococcal nuclease from Sigma, Ltd., Poole, UK, spleen phosphodiesterase, pancreatic DNase and venom phosphodiesterase from Worthington (Millipore (UK) Ltd.), London.

M13 17-mer primer and hybridisation probe primer were from New England Biolabs Inc. Beverly, MA, cytidylyl-(3′-5′)-uridine (CpU) from Sigma Ltd., Poole, UK, deoxynucleoside triphosphates (dNTPs) and dideoxynucleoside triphosphates (ddNTPs) from P-L Biochemicals Inc. Milwaukee, WI.

DEAE-Sephadex (A50) was from Pharmacia Inc., Uppsala, Sweden, DEAE-cellulose (DE52), phosphocellulose (P11) and AE-cellulose paper (AE81) from Whatman Biochemicals Ltd., Maidstone, UK.

L-[$^{35}$S]methionine (sp. act. >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 plc, Amersham, UK. Deoxyadenosine 5′-[α-$^{35}$S]thiotriphosphate (18.5 TBq/nmol) was from New England Nuclear, Dreieich, FRG.

*DNA; preparation, analysis and cloning*

Phage λ and T4 were prepared, purified and DNA extracted as described by Wilson *et al.* (1977); M13 RF DNA was prepared by the method of Messing (1983). For λCM21 DNA it was necessary to induce a lysogen, C600 (λCM21), and concentrate the cells before lysing with chloroform (see Arber *et al.*, 1983). DNA was digested with enzymes using the buffers recommended by the suppliers. Fragments were separated in 0.8% (w/v) agarose gels in TBE buffer, pH 8.2 (Biggin *et al.*, 1983), and isolated from preparative gels by electroelution in dialysis tubing followed by passage through DEAE-cellulose (Maniatis *et al.*, 1982). Small fragments were isolated from acrylamide gels (Maxam and Gilbert, 1977). Sonicated fragments were generated from circular DNA (Deininger, 1983); those in the size range 300–600 bp were incubated with Klenow polymerase and dNTPs to repair ends and ligated to M13 vector cut at the *Sma*I site. All ligations used T4 DNA ligase.

*Recovery of recombinants*

λ DNA was packaged as described by Hohn in Arber *et al.* (1983). Competent cells were generally made by the method of Lederberg and Cohen (1974), but the efficiency of recovery of M13 recombinants was increased by the modification of Hanahan (1983).

*Labelling of DNA in vitro, hybridisation and autoradiography*

Double-stranded DNA probes were labelled as described by Rigby *et al.* (1977). Single-stranded probes were made from a template of M13 recombinant DNA by the method of Hu and Messing (1982). DNA for hybridisation was transferred to nitrocellulose filters from plaques (Benton and Davis, 1977) or from cleared lysates (Hu and Messing, 1982).

*DNA sequencing of M13 recombinants*

Single-stranded template DNA was prepared by the method of Sanger *et al.* (1980), sequenced by the dideoxy chain-termination method using [α-$^{35}$S]thio-dATP, and analysed on buffer gradient acrylamide gels (Biggin *et al.*, 1983). The synthetic oligonucleotides used as primers were a 17-mer, 5′ AAACAGTTCATCCTAAT3′ and a 15-mer 5′ GTTCGACAAAGCCTC3′.

*Analysis of polypeptides following infection of u.v. light irradiated cells*

Protein synthesis was detected by pulse-labelling with L-[$^{35}$S]methoinine (Jaskunas *et al.*, 1975) after phage infection of u.v.-irradiated M159, and samples were analysed on 15% linear polyacrylamide gels containing SDS; these were made from a stock solution of 44% acrylamide, and 0.3% bis-acrylamide to improve separation of polypeptides (Hancock *et al.*, 1976).

*Infected cells for enzyme assays*

Infections of *E. coli* cultures by either phage T4 or λ, and induction of lysogens were carried out as described by Murray *et al.* (1979). For phages without amber mutations, the cells were recovered by centrifugation 10–15 min after infection.

*Assays for 5′-kinase and 3′-phosphatase activity*

The assay for 5′-kinase activity was based on the conversion of a dinucleoside monophosphate (CpU) to the dinucleotide (pCpU) (Murray, 1973). This quick, simple assay is ideal for large numbers of samples, but in crude extracts the kinase activity may be manifest as the appearance of radioactive pC resulting from the action of nuclease upon the primary product pCpU (Figure 6). Cell pellets were resuspended in 0.4 ml 50 mM Tris-HCl pH 7.4, 1 mM 2-mercaptoethanol in snap cap tubes and sonicated in six bursts of 3 s, ensuring that the tubes remain cool throughout. After centrifuging to remove cell debris, the supernatant (2.5 μl) 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, 1 μCi (37 Bq) [γ-$^{32}$P]ATP (sp. act. 110 TBq/nmol) in a 10 μl volume at 37°C for 30 min. Products were analysed by ionophoresis on AE-cellulose paper at pH 3.5 (Murray, 1973).

5′-Hydroxyl oligonucleotides were obtained by digestion of salmon sperm DNA with pancreatic DNase (DNase I) followed by alkaline phosphatase, or digestion of yeast tRNA with RNase T1.

5′-[$^{32}$P]dCMP and 3′-[$^{32}$P]dCMP, the substrates for 3′-phosphatase assays, were isolated by paper ionophoresis of enzymic digests of nick-translated λ DNA. 5′-[$^{32}$P]dCMP was obtained from sequential digestion with pancreatic DNase (DNase I) and venom phosphodiesterase, while 3′-[$^{32}$P]dCMP was isolated from sequential digestion with micrococcal nuclease and spleen phosphodiesterase. The mononucleotides were eluted from AE-cellulose paper with 30% (v/v) triethylamine carbonate solution (Murray, 1970). The assay for 3′-phosphatase activity was described by Becker and Hurwitz (1967); products were analysed by paper ionophoresis.

*Purification of polynucleotide kinase*

ED8689 cells at $OD_{650}$ 0.4 were infected with λ CM8 at a multiplicity of 2 (the highest possible with our phage lysate). After 10 min at 37°C cells were harvested and frozen at −20°C.

The enzyme was prepared from extracts essentially by the procedure described by Richardson (1965), except that DEAE-Sephadex was used in place of the DEAE-cellulose fractionation. The DEAE-Sephadex A50 column was equilibrated with 0.01 M potassium phosphate buffer pH 7.5 containing 0.01 M 2-mercaptoethanol, and washed with the same buffer after loading. Protein was eluted with equilibration buffer containing 0.1 M KCl, 5 ml fractions were collected and assayed for 5′-kinase activity as described above: active fractions were pooled and fractionated on phosphocellulose as described by Richardson (1965).

The defective polypeptide was purified in the same way but from cells infected with NM1108 and harvested 2 h after infection, since NM1108 is defective in lysis.

## Acknowledgements

We thank our colleagues for their help, particularly K.Murray for sharing his expertise. We are also indebted to E.Kawashima (Biogen, Geneva) and J.Keyte (Biochemistry, Leicester) for synthetic oligonucleotides; B.Kutter and L.Snyder for phage strains; B.Kutter and K.Rand for exchange of unpublished information; the Cambridge MRC Laboratory of Molecular Biology for initial help with DNA sequencing techniques; L.Sawyer for analysis of polypeptide structures; D.J.Finnegan, W.Loenen and K.Murray for critical comments on the manuscript, E.McCready and S.Dickson for typing and A.Wilson for figures. C.A.Midgley was supported by a Science and Engineering Research Council studentship and the research by a grant from the Medical Research Council.

## References

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., Enquist,L., Hohn,B., Murray,N.E. and Murray,K. (1983) in

Hendrix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A. (eds.), *Lambda II*, Cold Spring Harbor Laboratory Press, NY, USA, pp. 433-466.
Becker,A. and Hurwitz,J. (1967) *J. Biol. Chem.*, **242**, 936-950.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963-3965.
Bolivar,F. (1978) *Gene*, **4**, 121-136.
Cameron,V. and Uhlenbeck,O.C. (1977) *Biochemistry (Wash.)*, **16**, 5120-5126.
Chu,F.K., Maley,G.F., Maley,F. and Belfort,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3049-3053.
David,M., Borasio,G.D. and Kaufmann,G. (1982a) *Proc. Natl. Acad. Sci. USA*, **79**, 7097-7101.
David,M., Borasio,G.D. and Kaufmann,G. (1982b) *Virology*, **123**, 480-483.
Deininger,P.L. (1983) *Anal. Biochem.*, **129**, 216-233.
Depew,R.E. and Cozzarelli,N.R. (1974) *J. Virol.*, **13**, 888-897.
Frischauf,A.-M., Lehrach,H., Poustka,A. and Murray,N.E. (1983) *J. Mol. Biol.*, **170**, 827-842.
Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1-19.
Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557-580.
Hancock,R.E.W., Hankte,K. and Braun,V. (1976) *J. Bacteriol.*, **127**, 1370-1375.
Hu,N. and Messing,J. (1982) *Gene*, **17**, 271-277.
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*, **257**, 458-462.
Kelley,W.S., Chalmers,K. and Murray,N.E. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5632-5636.
Kutter,E., Drivdahl,R. and Rand,K. (1984) *Genetics*, **108**, 291-304.
Lederberg,E.M. and Cohen,S.N. (1974) *J. Bacteriol.*, **119**, 1072-1074.
Lillehaug,J.R. (1977) *Eur. J. Biochem.*, **73**, 499-506.
Lowry,O.H., Rosebrough,N.J., Farr,D.L. and Randall,R.J. (1951) *J. Biol. Chem.*, **193**, 265-275.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY, USA.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20-78.
Mileham,A.J., Revel,H.R. and Murray,N.E. (1980) *Mol. Gen. Genet.*, **179**, 227-239.
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 Press, NY, USA, 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.
Novogrodsky,A. and Hurwitz,J. (1966) *J. Biol. Chem.*, **241**, 2923-2932.
Pai,E.G., Sachsenheimer,W., Schirmer,R.H. and Schultz,G.E. (1977) *J. Mol. Biol.*, **114**, 37-45.
Panet,A., van de Sande,J.H., Loewen,P.C., Khorana,H.G., Raae,A.J., Lillehaug,J.R. and Kleppe,K. (1973) *Biochemistry (Wash.)*, **12**, 5045-5050.
Rand,K.N. and Gait,M.J. (1984) *EMBO J.*, **3**, 347-402.
Richardson,C.C. (1965) *Proc. Natl. Acad. Sci. USA*, **54**, 58-165.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.
Runnels,J.M., Soltis,D., Hey,T. and Snyder,L. (1982) *J. Mol. Biol.*, **154**, 273-286.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.M. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.
Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F. and Peterson,G.B. (1982) *J. Mol. Biol.*, **162**, 729-773.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342-1346.
Sirotkin,K., Cooley,W., Runnels,J. and Snyder,L. (1978) *J. Mol. Biol.*, **123**, 221-233.
Snopek,T.J., Wood,W.B., Conley,M.P., Chen,P. and Cozzarelli,N.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 3355-3359.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 4731-4751.
Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945-951.
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.
Wood,W.B. and Revel,H.R. (1976) *Bacteriol. Rev.*, **40**, 847-868.

*Received on 8 July 1985*