The EMBO Journal vol.3 no.2 pp.397 – 402, 1984

Sequence and cloning of bacteriophage T4 gene 63 encoding RNA ligase and tail fibre attachment activities

K.N.Rand* and M.J.Gait

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

*To whom reprint requests should be sent

Communicated by S.Brenner

The sequence of gene 63 of bacteriophage T4 was determined by a shotgun approach. Small DNA fragments, derived by sonication of a restriction fragment that encompasses the region of gene 63, were cloned in M13 vectors and sequenced by the 'dideoxy' method. The position of the gene was established by comparison with the sequence of a gene 63 amber mutant. Knowledge of the DNA sequence of gene 63 and surrounding regions has allowed the construction of a clone of gene 63 in which RNA ligase production is under the control of the *lac* promoter of bacteriophage M13mp8. Infected *E. coli* cells can be induced to produce a protein indistinguishable from commercially available RNA ligase.

Key words: cloning/DNA sequence/RNA ligase/T4 gene 63

Introduction

Bacteriophage T4 RNA ligase was discovered in 1972 as an activity that catalyses the circularisation of homopolyribonucleotides containing a 3′-terminal hydroxyl group and a 5′-terminal phosphate to form a 3′ → 5′ phosphodiester bond (Silber *et al.*, 1972). The enzyme was later shown to carry out intermolecular joining reactions with either RNA (Walker *et al.*, 1975) or DNA (Sugino *et al.*, 1977; Moseman McCoy and Gumport, 1980). It is used in the synthesis of defined sequence oligoribonucleotides and in the 3′-labelling of RNA using a nucleoside-3′,5′-diphosphate (Bruce and Uhlenbeck, 1978). These and other uses of T4 RNA ligase have been recently reviewed (Uhlenbeck and Gumport, 1982).

In 1977 it was shown that RNA ligase is the product of gene 63 (gp63) of bacteriophage T4 (Snopek *et al.*, 1977). The same product was known to promote tail fibre attachment (TFA) to the baseplate of T4 (Wood and Henninger, 1969). These two activities of gp63 appear to be unrelated since they have different requirements *in vitro* (Snopek *et al.*, 1977). Although the TFA activity of the protein has a clear physiological role in phage assembly, the *in vivo* function of RNA ligase is less certain. One model proposes a role in DNA metabolism (Runnels *et al.*, 1982). Another suggests an involvement in processing of host tRNAs (David *et al.*, 1979, 1982).

For many applications involving RNA, the purification routes for RNA ligase are adequate, although lengthy, and make use of *Escherichia coli* infected with a replication-deficient T4 mutant that hyperproduces RNA ligase (Last and Anderson, 1976; Higgins *et al.*, 1977; Moseman McCoy *et al.*, 1979; Sugiura *et al.*, 1979; Hu *et al.*, 1982). However, the removal of traces of nucleases that remain after apparent homogeneity is attained is essential for some DNA joining applications, and the purification procedure is more laborious (reviewed by Uhlenbeck and Gumport, 1982). The cloning of gene 63 (g63) using either a plasmid or phage vector and a suitable host might allow substantial overproduction and enable easier and more reliable isolation of the enzyme, as has been obtained for T4 DNA ligase (Wilson and Murray, 1979; Murray *et al.*, 1979). Further studies of the interesting enzyme activities will be aided by the ability to encode and produce altered proteins via manipulation of a cloned version of g63. The ability to prepare large quantities of pure enzyme may allow for its crystallization and X-ray structural determination.

Previous attempts by us and others (S.Brown and W.Wood, personal communication) to obtain a clone of the whole of g63 have been unsuccessful. Mileham *et al.* (1980) were unable to clone an 11.5-kb *Hind*III fragment of T4 DNA known to contain the gene. After *Eco*RI digestion of a 9-kb *Hind*III/*Pst*I subfragment, all the resultant fragments could be cloned in bacteriophage lambda except one of ~1.3 kb. The position of this fragment with respect to the genetic map and the failure to obtain the marker rescue of a g63 amber mutant with any of the lambda recombinants led Mileham *et al.* to conclude that the 1.3-kb *Eco*RI fragment might contain g63.

Presumably either gp63 itself or some other structure or function of T4 in this region is deleterious to *E. coli* or to the vector used and has prevented the cloning of g63. We now describe an approach which has overcome this problem. Small (200 – 600 bp) fragments of T4 DNA, derived by sonication of a large restriction fragment known to contain g63, were cloned into M13 and sequenced by the 'dideoxy' method. Because cuts due to sonication occur essentially at random (Deininger, 1983) we hoped that some would occur within, and thus inactivate, whichever region is responsible for the cloning failures. By obtaining a sufficient number of overlaps we were indeed able to determine the complete sequence of g63 and surrounding regions. From this information we have been able to devise a way of constructing a full length clone of g63.

Results

Determination of the DNA sequence of g63

An 8.2-kb *Bgl*II fragment which is large enough to be sure of containing g63 migrates in an agarose gel to a well resolved position (O'Farrell *et al.*, 1980). [In later work a 9-kb *Hind*III/*Pst*I fragment covering the same region (Mileham *et al.*, 1980) was used.] This fragment was isolated and sheared by sonication. Fragments in the size range of 200 – 600 bp were eluted from an agarose gel, cloned into the *Sma*I site of M13mp8 and sequenced by the dideoxy method. Sequence data obtained to date include a stretch of ~4 kb which contains an open reading frame long enough to be capable of encoding RNA ligase. The position of this, in relation to various restriction enzyme sites, is shown in Figure 1. The position of the open reading frame corresponds well to the location of

© IRL Press Limited, Oxford, England.

g63 determined by genetic mapping (Wood and Revel, 1976; O'Farrell *et al.*, 1980).

Eighteen clones cover the region of the open reading frame and sequence information was obtained from both strands of DNA except for 137 bp at its 3′ end, a region from which it is difficult to obtain clones. The DNA sequence and the derived amino acid sequence of the possible product are shown in Figure 2. Use of a program (Staden, 1984) that uses the weight matrix of Stormo *et al.* (1982) suggested that the first ATG in the open reading frame is by far the most likely initiation codon. There is a probable ribosome binding site 8–12 nucleotides upstream of this ATG (Shine and Dalgarno, 1974). A protein encoded by this open reading frame would have 375 amino acids and a mol. wt. of 43 510 daltons. This lies within the range of observed values for RNA ligase, which are 41 000–45 000 by SDS-polyacrylamide gel electrophoresis (Snopek *et al.*, 1976; Moseman McCoy *et al.*, 1979; Last and Anderson, 1976), 47 000 by gel filtration (Last and Anderson, 1976) and 48 200 by sedimentation equilibrium ultracentrifugation (Last and Anderson, 1976).

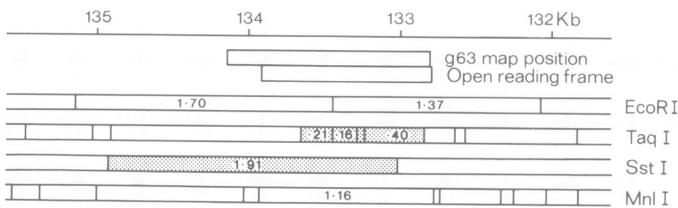

Fig. 1. Restriction map derived from the DNA sequence in the region 132–136 kb of the T4 genetic map. The genetic map position of g63 is from O'Farrell *et al.* (1980). The positions of the longest open reading frame and the recognition sites of enzymes most relevant to the work described here were aligned with the genetic map by reference to the *Eco*RI site that separates the 1.7- and 1.37-kb *Eco*RI fragments.

To confirm that this is g63, we cloned in M13mp8 and partly sequenced certain *Taq*I fragments from a total digest of the DNA of a g63 amber mutant, E1072. Clones containing the *Taq* fragments shaded in Figure 1 were detected by dot hybridisation (Hu and Messing, 1982) using probes made from clones used previously in the sequencing of this region. The sequence underlined in Figure 2 was determined in this way for the g63 amber mutant; the only difference between the two sequences is a C to T transition at position 394 that generates an amber codon.

Codon usage in g63

Certain codons that are rarely used in strongly expressed genes are used more frequently in g63; apart from AAU, these codons are the ones recognised by T4-encoded tRNAs (within boxes in Table I).

Construction of the g63 clone KR72

As can be seen from Figure 3, S205 was the only clone containing the 3′ end of g63 that we had been able to obtain and so was chosen as the starting point for construction of a g63

```
                   M  Q  E  L  F  N  N  L  M  E  L  C  K  D  S  Q  R  K  F  F  Y  S  D  D  V  S  A  S  G  R  T  Y  R  I
CACAAAAGAGGATACACAATGCAAGAACTTTTTAACAATTTAATGGAACTATGTAAGGATTCGCAGCGTAAGTTTTTTTACTCAGATGATGTAAGTGCATCTGGAAGAACTTACAGAATT
         10        20        30        40        50        60        70        80        90       100       110       120

 F  S  Y  N  Y  A  S  Y  S  D  W  L  L  P  D  A  L  E  C  R  G  I  M  F  E  M  D  G  E  K  P  V  R  I  A  S  R  P  M  E
TTCTCATATAATTATGCATCTTATTCTGATTGGTTACTTCCAGATGCACTAGAATGTCGTGGAATTATGTTTGAAATGGATGGAGAAAAACCAGTAAGAATTGCTTCTCGTCCTATGGAA
       130       140       150       160       170       180       190       200       210       220       230       240

 K  F  F  N  L  N  E  N  P  F  T  M  N  I  D  L  N  D  V  D  Y  I  L  T  K  E  D  G  S  L  V  S  T  Y  L  D  G  D  E  I
AAGTTTTTTAACTTGAATGAAAATCCGTTCACGATGAATATCGATTTAAACGATGTTGATTATATTCTAACAAAAGAAGACGGGTCTTTGGTATCAACTTATTTAGACGGTGATGAAATT
       250       260       270       280 Taq I 290       300       310       320       330       340       350       360

 L  F  K  S  K  G  S  I  K  S  E  Q  A  L  M  A  N  G  I  L  M  N  I  N  H  H  R  L  R  D  R  L  K  E  L  A  E  D  G  F
CTGTTCAAATCAAAGGGTTCAATCAAATCTGAGCAGGCTTTAATGGCTAATGGAATTTTGATGAATATTAATCACCATCGGTTGCGTGATAGACTTAAAGAATTAGCTGAAGATGGATTT
       370       380       390       400       410       420       430       440       450       460       470       480

 T  A  N  F  E  F  V  A  P  T  N  R  I  V  L  A  Y  Q  E  M  K  I  I  L  L  N  V  R  E  N  E  T  G  E  Y  I  S  Y  D  D
ACTGCTAACTTCGAATTCGTTGCCCCGACGAATAGAATCGTTCTTGCTTATCAAGAGATGAAAATTATTTTACTGAATGTTCGTGAAAACGAAACGGGTGAATACATTTCATACGATGAT
       490 Taq I EcoR I      510       520       530       540       550       560       570       580       590       600

 I  Y  K  D  A  T  L  R  P  Y  L  V  E  R  Y  E  I  D  S  P  K  W  I  E  E  A  K  N  A  E  N  I  E  G  Y  V  A  V  M  K
ATTTATAAAGATGCTACTCTTCGTCCGTATCTAGTTGAACGATACGAAATCGATAGCCCTAAATGGATAGAAGAAGCTAAAAATGCAGAAAACATCGAAGGCTATGTTGCTGTGATGAAA
       610       620       630       640     Taq I     660       670       680       690   Taq I             710       720

 D  G  S  H  F  K  I  K  S  D  W  Y  V  S  L  H  S  T  K  S  S  L  D  N  P  E  K  L  F  K  T  I  I  D  G  A  S  D  D  L
GATGGTTCTCATTTTAAAATTAAGTCTGACTGGTACGTGTCTCTTCATAGTACAAAAAGTTCATTAGATAATCCAGAAAAATTGTTTAAGACTATTATTGATGGTGCATCAGATGATCTT
       730       740       750       760       770       780       790       800       810       820       830       840

 K  A  M  Y  A  D  D  E  Y  S  Y  R  K  I  E  A  F  E  T  T  Y  L  K  Y  L  D  R  A  L  F  L  V  L  D  C  H  N  K  H  C
AAAGCAATGTATGCTGACGATGAATATTCATACAGAAAAATTGAAGCATTTGAAACGACTTATCTGAAGTACTTAGACCGAGCTCTGTTTTTAGTTCTTGACTGTCATAATAAGCATTGC
       850       860       870       880       890       900       910       920 Sst I 930       940       950       960

 G  K  D  R  K  T  Y  A  M  E  A  Q  G  V  A  K  G  A  G  M  D  H  L  F  G  I  I  M  S  L  Y  Q  G  Y  D  S  Q  E  K  V
GGTAAGGATAGAAAGACTTATGCAATGGAAGCACAAGGTGTTGCTAAAGGTGCTGGAATGGATCACCTGTTCGGTATCATCATGAGCTTATACCAGGGGTACGATAGTCAAGAAAAGGTC
       970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080

 M  C  E  I  E  Q  N  F  L  K  N  Y  K  K  F  I  P  E  G  Y  *
ATGTGTGAAATCGAACAGAATTTTTTGAAAAATTATAAAAAATTTATCCCAGAAGGATACTAAGTTGTTTACAAGTCCCTCGTGTTGT
      1090 Taq I    1100      1110      1120      1130      1140      1150      1160
```

Fig. 2. DNA and protein sequence of g63. The region that was also sequenced from the g63 amber mutant E1072 is underlined; an arrow at position 394 indicates where a T was found in the E1072 sequence. The DNA sequence shown is of the expected fragment after cutting with *Mnl*I (Roberts, 1983).

clone. The S205 insert includes the unique *Sst*I site within g63. A 560-bp *Hpa*II fragment containing the 285-bp T4 insert was cut from S205 and inserted into the *Acc*I site of pUC9 to form pUC9H. A 1.91-kb *Sst*I fragment from T4 DNA (see Figure 1) that contains the 5′ part of g63 was cloned into the *Sst*I site of M13mp12 to purify this region of the T4 genome and then recloned into the *Sst*I site (near the 3′ end of g63) of pUC9H to make PUC9HS and reform g63. KR72 was obtained by cloning into the *Sma*I site of M13mp8 a 1.1-kb *Mnl*I fragment (containing g63, see Figure 1) from pUC9HS. KR72 contains the whole of g63, orientated such that it can be expressed from the *lac* promoter of M13mp8, plus 18 base pairs upstream (including the Shine-Dalgarno sequence) and 25 base pairs downstream of the gene. We have checked by DNA sequencing that no alteration in sequence occurred during the construction at either the beginning or at the end of the gene.

Expression of cloned gene 63 in E. coli

KR72 contains within the *Sma*I site of M13mp8 the T4 *Mnl*I fragment, the sequence of which is shown in Figure 2. A protein with the same mobility on SDS polyacrylamide gels as a

Table I. Codon usage: comparison between g63 and strongly expressed genes of *E. coli*

	U					C					A					G					
	aa	*coli*	%	g63	%	aa	*coli*	%	g63	%	aa	*coli*	%	*g63*	%	*aa*	*coli*	%	g63	%	
U	F	39	26	13	68	S	93	36	9	36	Y	34	26	14	56	C	13	36	4	80	U
	F	113	74	6	32	S	87	34	0	0	Y	98	74	11	44	C	23	64	1	20	C
	L	12	3	11	32	S	6	2	9	36	–	–	–	–	–	–	–	–	–		A
	L	16	4	6	18	S	12	5	1	4	–	–	–	–	–	W	25	100	3	100	G
C	L	26	6	8	24	P	21	10	2	22	H	19	20	5	71	R	223	67	6	38	U
	L	33	8	0	0	P	2	1	0	0	H	75	80	2	29	R	101	30	0	0	C
	L	3	1	4	12	P	26	12	4	44	Q	38	18	4	50	R	3	1	2	13	A
	L	345	79	5	15	P	162	77	3	33	Q	169	82	4	50	R	1	0.3	1	6	G
A	I	67	20	15	60	T	103	36	7	54	N	13	8	15	71	S	10	4	4	16	U
	I	262	79	9	36	T	137	48	0	0	N	159	92	6	29	S	49	19	2	8	C
	I	2	1	1	4	T	15	5	2	15	K	259	71	19	63	R	3	1	7	44	A
	M	140	100	15	100	T	28	10	4	31	K	106	29	11	37	R	1	0.3	0	0	G
G	V	192	44	8	57	A	173	37	13	57	D	116	36	25	81	G	226	54	9	47	U
	V	41	9	1	7	A	48	10	1	4	D	204	64	6	19	G	174	42	1	5	C
	V	119	27	3	21	A	119	25	9	39	E	333	76	30	94	G	4	1	7	37	A
	V	83	19	2	14	A	129	28	0	0	E	106	24	2	6	G	14	3	2	11	G

The data for strongly expressed genes are taken from Grosjean and Fiers (1982). Data are normalised in the '%' column to take into account the different amino acid compositions of the gene products; in this column the number of occurrences of a particular codon is expressed as a percentage of the total number of codons corresponding to the same amino acid. The codons in boxes are those that are complementary to the anticodons of seven T4-encoded tRNAs (Fukada and Abelson, 1980). The sequence of the anticodon of an eighth T4 tRNA is uncertain (Fukada and Abelson, 1980).

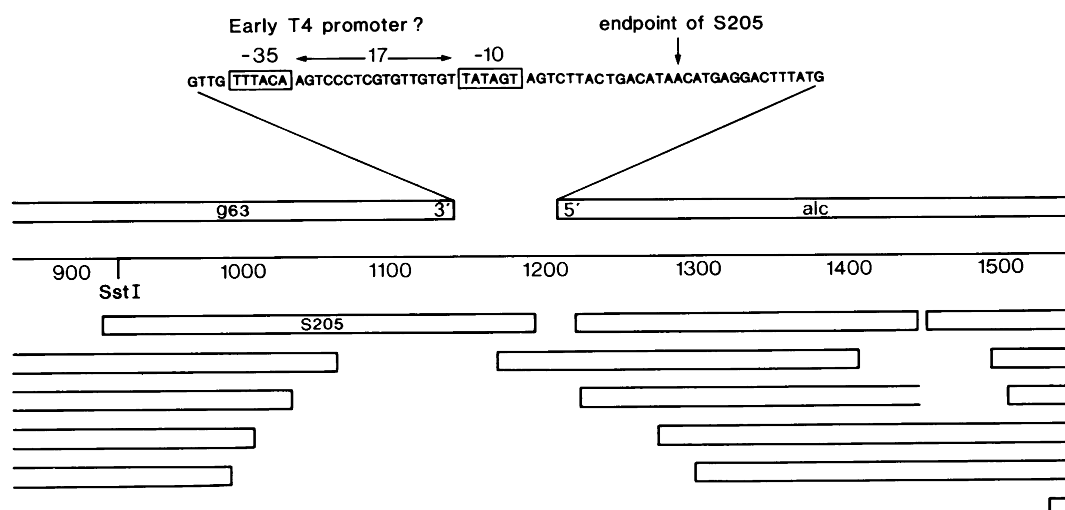

Fig. 3. DNA sequence between gene 63 and the gene *alc* (Snyder *et al.*, 1976). The position of a possible 'early' promoter (which could be recognised by unmodified RNA polymerase) is shown. Extents of the individual sequence readings are shown by horizontal bars; where endpoints of the inserts are known they are depicted by vertical lines. It is possible that a mutated version of the wild-type T4 sequence is shown here, for we have only been able to obtain one clone (S205) containing the putative promoter.

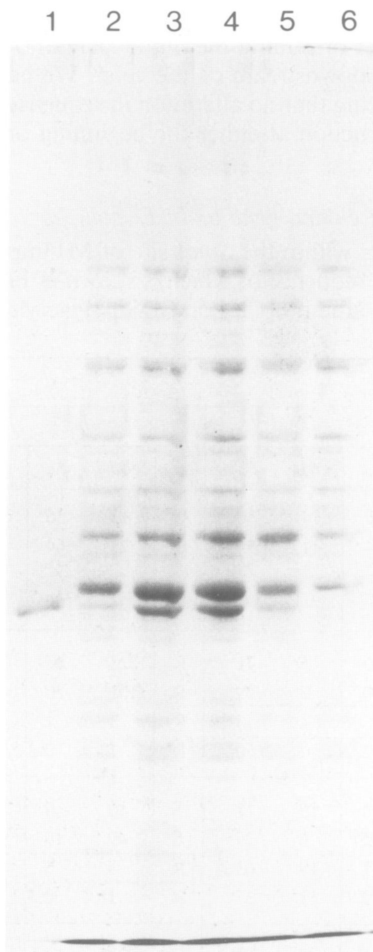

Fig. 4. SDS 10% polyacrylamide gel (Coomassie Blue stained) of phage infected cell lysates: **slot 1**, commercial RNA ligase (P.-L. Biochemicals) standard; **slots 2–4**, KR72 infection, 1, 2 and 3 h after induction by IPTG; **slot 5**, KR72 infection, same growth time as **slot 4**, uninduced; **slot 6**, M13mp8 infection, 3 h after induction by IPTG.

commercially available sample of RNA ligase is present in crude extracts of KR72-infected cells, but not in *E. coli* cells infected with M13mp8 (Figure 4). KR72-infected cells induced by isopropyl thiogalactoside (IPTG) showed a substantial increase in production of RNA ligase over uninduced cells and the enzyme accounted for ~5–10% of the protein extracted as judged by scanning densitometry. Crude cell-free extracts were diluted to 0.8 mg/ml of protein (measured with the Biorad protein assay kit) and assayed for RNA ligase by the method of Bruce and Uhlenbeck (1978). To take into account the low level of background activity, the apparent activity in similar extracts of uninfected cells was subtracted for each assay. IPTG-induced KR72-infected cells produce ~500 units of RNA ligase per mg of extracted protein. IPTG is a gratuitous inducer of the *lac* operon, and *o*-nitrophenyl fucopyranoside (ONPF) is a competitive inhibitor of induction (Jayaraman *et al.*, 1966).

Another indication that the expressed product from KR72-infected *E. coli* cells is active is its ability to react with ATP to form an adenylylated enzyme, the first step in the pathway of RNA ligase catalysis. Adenylylated and unadenylylated RNA ligase can be separated on an analytical level by SDS-polyacrylamide gel electrophoresis (Higgins *et al.*, 1977), and it can be seen from Figure 5 that, whereas crude lysates contain

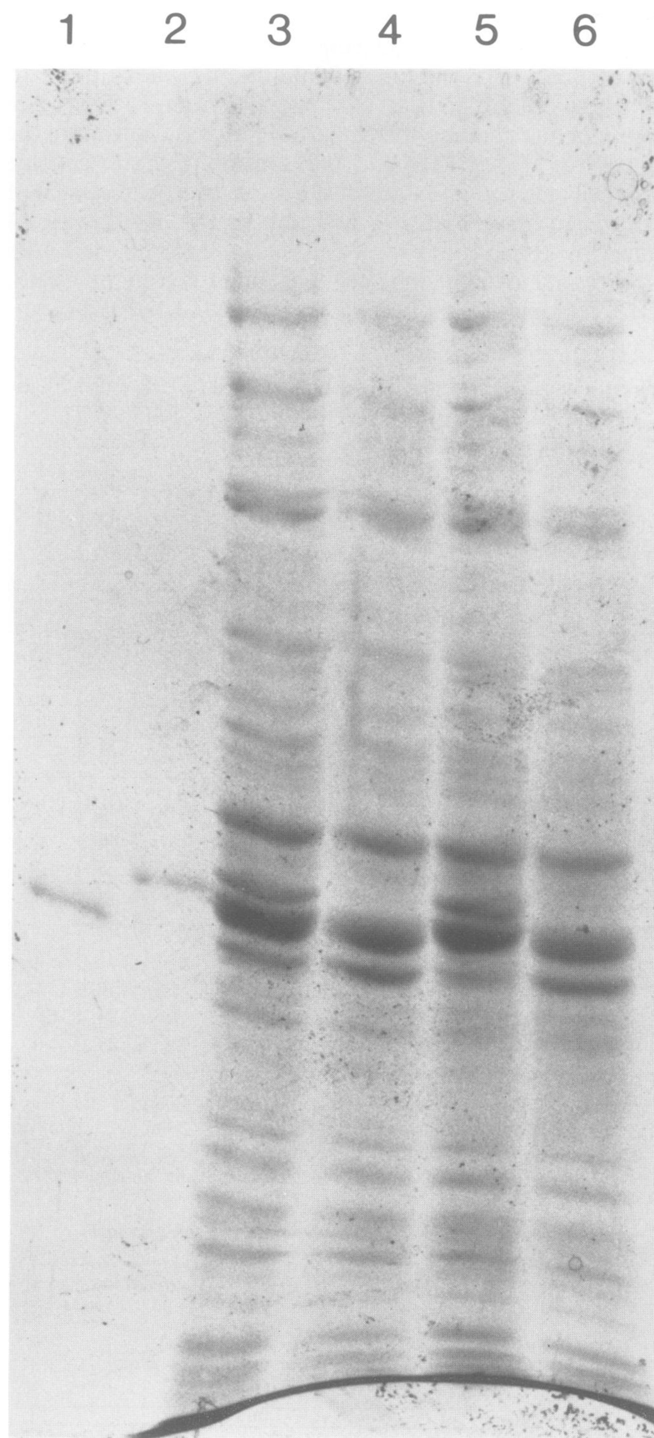

Fig. 5. SDS 10% polyacrylamide gel (Coomassie Blue stained) of adenylylation and de-adenylylation reactions (see Materials and methods). **Slot 1**, commercial RNA ligase; **slot 2**, commercial RNA ligase after ATP treatment; **slot 3**, KR72-infected cell lysate; **slot 4**, KR72 lysate after pyrophosphate treatment; **slot 5**, KR72 lysate after ATP treatment; **slot 6**, KR72 lysate after first ATP treatment and then pyrophosphate treatment.

varying amounts of adenylylated and free enzyme (cf. Figures 4 and 5), in the presence of sodium pyrophosphate essentially only free enzyme is seen. Moreover, addition of ATP to the crude lysate results in complete conversion to adenylylated enzyme and subsequent addition of pyrophosphate once again converts the enzyme into unadenylylated form.

Discussion

By the cloning and sequencing of small DNA fragments generated by random shearing we have obtained the sequence of a region that we and others had previously been unable to clone in its entirety. Examination of this sequence and use of an analysis program based on the data of Hawley and McClure (1983) suggest that there is a promoter located immediately 3′ to g63 (see Figure 3). A strong T4 'early' promoter recognised by RNA polymerase *in vitro* maps in this region (H.Gram and W.Rüger, personal communication). The promoter is within the 1.37-kb *Eco*RI fragment that we and others have been unable to clone, and in a small region from which it was difficult to obtain clones for the sequencing. Efficient promoters can interfere with plasmid replication (Stueber and Bujard, 1982) and it is well known that strong promoters are difficult to clone. However, one clone described here, S205, appears to contain the T4 promoter. One possibility is that a mutation in S205 reduces the activity of the promoter. Another is that it is the expression of the early part of the T4 gene downstream of the promoter that is lethal to the host or vector. Our limited data are consistent with the second possibility; we have been unable to isolate clones which contain both the 5′ end of this gene and the T4 promoter, or clones containing the 5′ end orientated such that its expression could occur via the *lac* promoter of M13mp8 (see Figure 3). The gene is *alc* (E.Kutter, R.Drivdahl and K.Rand, in preparation) whose product in some way causes the blocking of host transcription and the unfolding of the host chromosome (Sirotkin *et al.*, 1977). It should not be surprising, therefore, to be unable to isolate any clone in which *alc* is expressed. That expression of even just the 5′ part of *alc* is deleterious is indicated by the fact that an *alc*$^-$ mutant (*alc*10) was used in this work, and that most of the DNA fragments used in the sequencing are too small to encode the whole of the gene.

Whatever the cause of our failure to clone the region containing g63, it is clear that the clone S205 either lacks it or has a mutated version of it. As it was the only clone containing the 3′ end of g63 that we had been able to obtain after numerous cloning experiments, it was chosen as the starting point in the construction of a full length g63 clone. Reference to the DNA sequence thus far obtained from the g63 region allowed us to choose a suitable restriction fragment carrying the 5′ region of g63 to use in the first part of the construction (the 1.91-kb *Sst*I fragment, see Figure 1). *Mnl*I was then used to remove most of the T4 sequence surrounding g63 to form KR72. This clone contains within the *Sma*I site of M13mp8, the T4 *Mnl*I fragment whose sequence is shown in Figure 2.

Upon infection of *E. coli* cells with KR72 and induction of the *lac* promoter, RNA ligase is produced at ~5–10% of the soluble protein. The standard method of assay of RNA ligase activity entails determining the circularisation of [5′-^{32}P]rA_n through the conversion of the label to a diester bond that is resistant to phosphomonoesterase (Silber *et al.*, 1972). However it is difficult to quantify accurately the amount of enzyme with this assay in crude extracts (Brennan *et al.*, 1983). An alternative method of assay of RNA ligase involves the measurement of the addition of [5′-^{32}P]pCp to tRNA (Bruce and Uhlenbeck, 1978). We have found that there is an apparent reduction of activity of commercially prepared RNA ligase when added to a crude cell lysate of M13-infected *E. coli* cells. However, assuming that the activities of both *in situ* and added RNA ligase are equally reduced in this situation, we estimate that ~500 units/mg of extracted protein are produced in KR72-infected cells after induction with IPTG. This level is roughly consistent with the amount of gene 63 product seen by SDS polyacrylamide gel electrophoresis, and this suggests that the expressed product is fully active. This is further corroborated by the ability of the product to be adenylylated and deadenylylated without difficulty.

Table II. Induction of RNA ligase in KR72-infected cells

Addition (after 2 h)	Relative activity of RNA ligase
None	29
1 mM Isopropyl-β-D-thiogalactoside (IPTG)	100
1 mM *o*-nitrophenyl-β-D-fucopyranoside (ONPF)	0

As shown in Table II, RNA ligase is produced in significant amounts in the absence of added inducer. That this level of synthesis is due to inducer(s) in the 2 x TY medium is indicated by the fact that addition of ONPF a competitive inhibitor of induction (Jayaraman *et al.*, 1966) eliminates RNA ligase activity. Thus to reliably switch off all gene expression from the *lac* promoter on a multiple-copy vector such as M13mp8, a combination of the *lac*I^q mutation and ONPF might be needed.

We plan to investigate further the expression of the enzyme in *E. coli* using both this and other vector systems. We have shown in Table I that the codon usage of g63 differs from the data of Grosjean and Fiers (1982) for highly expressed genes in *E. coli*. It is possible that this could limit the level of production of active enzyme obtainable. However, a recent report by Young *et al.* (1983) suggests that rare codon usage may not be a limiting factor in obtaining high level synthesis of proteins in *E. coli*. In any event, now that g63 has been cloned it should be relatively straightforward to obtain reliable purification of nuclease-free RNA ligase.

Furthermore, expression of g63 in M13 vectors (for example, KR72) is particularly useful in that this system can be used both for mutagenesis of the gene and analysis of mutants by rapid DNA sequencing, and should aid our study of the gene 63 product and its TFA and RNA ligase activities.

Materials and methods

T4 strains and vectors

The T4 strain used for making cytosine-containing DNA was JW800 (*am*N55X5[42], *am*E51[56], *nd*28[*den*A], D1923[*den*B], *alc*10) supplied to us by H.Gram and W.Rüger. The gene 63 amber mutant strain E1072 was from the stocks of S.Brenner. M13mp8 was described by Messing and Vieira (1982). M13mp12 contains cloning sites for *Eco*RI, *Sma*I, *Xba*I, *Sst*I, *Xho*I, *Bam*HI and *Hind*III fragments. It was constructed by H.W.D.Matthes, J.Karn, M.J.Gait and S.Brenner (in preparation). pUC9 was described by Vieira and Messing (1982).

Enzymes

All restriction endonucleases used were from New England Biolabs apart from *Sst*I which was purchased from BRL. DNA polymerase (Klenow) and calf intestinal alkaline phosphatase were from Boehringer and T4 DNA polymerase and T4 RNA ligase were from P.-L. Biochemicals. T4 DNA ligase was a gift from David Bentley.

Preparation of cytosine-containing T4 DNA

The T4 *alc* mutant strain JW800 was grown first in *E. coli* K803 (Wood, 1966) and then in *E. coli* B834 GalU$^-$ (Runnels and Snyder, 1978) by the method of Snyder *et al.* (1976). Phage were precipitated with 6% polyethylene glycol (Yamamoto *et al.*, 1970) and further purified by CsCl centrifugation. After phenol extractions the DNA was ethanol precipitated, washed in 70% ethanol and redissolved in 10 mM Tris-Cl pH 7.4, 0.1 mM EDTA.

Isolation of plasmid DNA

This was done essentially as described by Birnboim and Doly 1979).

Dot hybridisation

The method of Hu and Messing (1982) was used. The radioactive M13 probes were made by the method of Brown *et al.* (1982).

Isolation of DNA fragments from gels

Two methods were used. In one, LGT agarose (Sea Plaque) was used for electrophoresis. Gel slices were melted at 65°C. Two volumes of 10 mM Tris-HCl (pH 7.4), 0.1 mM Na_2 EDTA were added and agarose was removed by three phenol extractions. Sodium acetate was added to 0.3 M and DNA was ethanol precipitated (sometimes after addition of carrier tRNA). The precipitate was washed in 70% ethanol before redissolving in 10 mM Tris-HCl (pH 7.4), 0.1 mM EDTA. In the second method HGT agarose (Seakem) was used, and DNA was electroeluted (McDonell *et al.*, 1977; Chouikh *et al.*, 1979).

DNA sequencing

The sonication, end-repair and size fractionation of DNA fragments were carried out by the methods of Deininger (1983) except that the sonicator was from Heat Systems-Ultrasonics, Inc. (Model W-375).

The fragments were ligated to M13mp8 replicative form DNA that had been cut with *Sma*I and treated with calf intestine alkaline phosphatase. Transfection was originally done by the method of Dagert and Ehrlich (1979) but in later work we followed the improved protocol of Hanahan (1983). Growth of clones was as described by Sanger *et al.* (1980) except that JM103 (Messing *et al.*, 1981) and in some cases CMK603 (a r_k^- host for M13, D.Bentley, personal communication) was used in place of JM101. Sequencing with chain-terminating inhibitors (Sanger *et al.*, 1977) and preparation of buffer gradient gels was carried out as described (for [α-^{32}P]dATP) by Biggin *et al.* (1983). DNA sequences were compiled and analysed using the computer programs of Staden (1982, 1984).

Expression of g63 in E. coli

An overnight culture of *E. coli* TG1 (an r_k^- derivative of JM101, T.Gibson, personal communication) was diluted 1:100 into 10 ml of 2 x TY medium (16 g Bactotryptone, 10 g yeast extract, 5 g sodium chloride in 1 l water adjusted to pH 7.4) and infected at high multiplicity with phage KR72 (or control phage M13mp8). After 2 h shaking at 37°C (A_{595} ~0.5) half of the culture was transferred to a separate tube and 25 μl IPTG (20 mg/ml) added. Both tubes were shaken for a further 3 h until the cells were close to saturation (A_{595} = 1.3 – 1.4). 1 ml aliquots of culture were centrifuged at low speed in plastic Eppendorf tubes and the pellets washed with 500 μl of 50 mM Tris-HCl pH 7.4, resuspended in 100 μl of the same buffer plus 1 mM EDTA, 10 mM β-mercaptoethanol and 0.1 mM phenylmethylsulphonylfluoride and frozen at –20°C overnight. After slow thawing to room temperature the cells were sonicated for 40 s in a Heat Systems sonicator (Model W-375), cooled on ice for 5 min and spun at high speed for 3 min. The cell lysates were analysed on an SDS polyacrylamide gel essentially by the method of Laemmli (1970). Gel tracks were scanned by a Camag flat bed densitometer.

Crude cell lysates were also assayed for RNA ligase activity by the method of Bruce and Uhlenbeck (1978) involving measurement of the joining of [5'-^{32}P]cytidine-3',5'-bisphosphate to tRNA. RNA ligase obtained from P.-L. Biochemicals was used as standard. One unit of activity is defined as the amount that catalyses the formation of 1 nmol of phosphatase-resistant ^{32}P in 30 min at 37°C with 5'-^{32}P-oligo(rA)$_n$ as substrate.

Adenylylation was carried out in sonication buffer at room temperature for 15 min by addition of magnesium chloride to 50 mM and ATP to 0.15 mM. De-adenylylation was carried out at room temperature for 10 min by addition of sodium pyrophosphate (pH 7.5) to 10 mM. In both cases reaction was terminated by addition of gel loading buffer/dye mix followed by heating to 100°C for 2 min.

Acknowledgements

We thank Sydney Brenner, Mark Biggin, Toby Gibson, Jon Karn, Rodger Staden and especially Keith McKenney and Greg Winter, for helpful advice, vectors and bacterial strains. We also thank W.Rüger for the gift of strain JW800 and Noreen Murray for JW819 and Dick Gumport for protocols for adenylylation of RNA ligase. K.N.R. gratefully acknowledges the award of an MRC Training Fellowship.

References

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

Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.

Brennan,C.A., Manthey,A.E. and Gumport,R.I. (1983) *Methods Enzymol.*, **100**, 38-52.

Brown,D.M., Frampton,J., Goelet,P. and Karn,J. (1982) *Gene*, **20**, 139-144.

Bruce,A.G. and Uhlenbeck,O.C. (1978) *Nucleic Acids Res.*, **5**, 3665-3677.

Chouikh,Y., Volovitch,M. and Yot,P. (1979) *Mol. Biol. Rep.*, **5**, 237-239.

Dagert,M. and Ehrlich,S.D. (1979) *Gene*, **6**, 23-28.

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

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

Deininger,P.L. (1983) *Anal. Biochem.*, **129**, 216-223.

Fukada,K. and Abelson,J. (1980) *J. Mol. Biol.*, **139**, 377-391.

Grosjean,H. and Fiers,W. (1982) *Gene*, **18**, 199-209.

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

Hawley,D.K. and McClure,W.R. (1983) *Nucleic Acids Res.*, **11**, 2237-2255.

Higgins,N.P., Geballe,A.P., Snopek,T.J., Sugino,A. and Cozzarelli,N.R. (1977) *Nucleic Acids Res.*, **4**, 3175-3186.

Hu,M., Wang,A. and Hua,H. (1982) *Anal. Biochem.*, **125**, 1-5.

Hu,N. and Messing,J. (1982) *Gene*, **17**, 271-277.

Jayaraman,K., Müller-Hill,B. and Rickenberg,H.V. (1966) *J. Mol. Biol.*, **18**, 339-343.

Laemmli,U.K. (1970) *Nature*, **227**, 680-685.

Last,J.A. and Anderson,W.F. (1976) *Arch. Biochem. Biophys.*, **174**, 167-176.

McDonell,M.W., Simon,M.N. and Studier,F.W. (1977) *J. Mol. Biol.*, **110**, 119-146.

Messing,J., Crea,R. and Seeburg,P.H. (1981) *Nucleic Acids Res.*, **9**, 309-321.

Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.

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

Moseman McCoy,M.I., Lubben,T.H. and Gumport,R.I. (1979) *Biochim. Biophys. Acta*, **562**, 149-161.

Moseman McCoy,M.I. and Gumport,R.I. (1980) *Biochemistry (Wash.)*, **19**, 635-642.

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

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

Roberts,R.J. (1983) *Nucleic Acids Res.*, **11**, r135-r167.

Runnels,J. and Snyder,L. (1978) *J. Virol.*, **27**, 815-818.

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

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.

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

Silber,R., Malathi,V.G. and Hurwitz,J. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 3009-3013.

Sirotkin,K., Wei,J. and Snyder,L. (1977) *Nature*, **265**, 28-32.

Snopek,T.J., Sugino,A., Agarwal,K. and Cozzarelli,N.R. (1976) *Biochem. Biophys. Res. Commun.*, **68**, 417-424.

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

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

Staden,R. (1984) *Nucleic Acids Res.*, in press.

Stormo,G.D., Schneider,T.D., Gold,L. and Ehrenfeucht,A. (1982) *Nucleic Acids Res.*, **10**, 2997-3011.

Stueber,D. and Bujard,H. (1982) *EMBO J.*, **1**, 1399-1404.

Sugino,A., Snopek,T.J. and Cozzarelli,N.R. (1977) *J. Biol. Chem.*, **252**, 1732-1738.

Sugiura,M., Suzuki,M., Ohtsuka,E., Nishikawa,S., Uemura,H. and Ikehara, M. (1979) *FEBS Lett.*, **97**, 73-76.

Uhlenbeck,O.C. and Gumport,R.I. (1982) *The Enzymes*, Vol. **XV**, published by Academic Press, Inc., NY, pp. 31-58.

Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.

Walker,G.C., Uhlenbeck,O.C., Bedows,E. and Gumport,R.I. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 122-126.

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

Wood,W.B. (1966) *J. Mol. Biol.*, **16**, 118-133.

Wood,W.B. and Henninger,M. (1969) *J. Mol. Biol.*, **39**, 603-618.

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

Yamamoto,K.R., Alberts,B.M., Benzinger,R., Lawhorne,L. and Treiber,G. (1970) *Virology*, **40**, 734-744.

Young,J.F., Desselberger,U., Palese,P., Ferguson,B., Shatzman,A.R. and Rosenberg,M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6105-6109.

Received on 4 November 1983; revised on 6 December 1983