Nucleic Acids Research, Vol. 20, No. 17 4591–4598

Construction and use of λ P_L promoter vectors for direct cloning and high level expression of PCR amplified DNA coding sequences

Xinbo Cheng and Thomas A.Patterson*
Biotechnology, DuPont Merck Pharmaceutical Company, Wilmington, DE 19880-0400, USA

Received April 27, 1992; Revised and Accepted July 22, 1992

ABSTRACT

A set of plasmid vectors which allow single-step cloning and expression of PCR-amplified DNA coding sequences has been constructed. The vectors contain the phage λ P_L promoter, a synthetic translation initiation region (TIR), and convenient cloning sites. The cloning sites provide all or part of an AUG translation initiation codon and facilitate the precise fusion of target DNA sequences to vector transcriptional and translational signals. The vectors were constructed with synthetic TIRs because there is evidence which suggests that the efficiency of the phage λ *c*II gene TIR present in the parental vector depends strongly on information contained within the *c*II N-terminal coding sequence. Bovine brain 14-3-3 η chain cDNA was PCR-amplified and used to demonstrate the expression capacity of the newly constructed vectors. A significant increase in expression of 14-3-3 protein was observed when synthetic TIRs were used in the place of the *c*II TIR. Expression levels vary from 15% to 48% of total cell protein. The effects of a reported translational enhancer from phage T7 on expression of the 14-3-3 protein are also discussed. The vectors should be generally useful for high level heterologous protein expression in *Escherichia coli*.

INTRODUCTION

Bacteriophage λ P_L promoter based vector systems have been shown to be very efficient and convenient for the expression of foreign genes in *E.coli* (1–4). The heat inducibility of the P_L promoter makes it especially attractive for large scale production and purification of gene products. One of the most frequently used vectors containing the λ P_L promoter is pJL6 (5). It has been used to direct high level expression of a number of proteins as fusions to the first 14 amino acids of the phage λ *c*II protein (*c*II leader) (1,4–6). We have successfully expressed human DNA polymerase β (7), yeast APN1 protein (8), and human apurinic/apyrimidinic endonuclease (HAP1h) (9) as *c*II fusion proteins using this system (unpublished data).

It is often desirable to express a protein with an authentic N-terminal sequence. However, previously published data (4) and our data indicate that removal of the *c*II leader coding sequence from pJL6-based fusion expression vectors can drastically decrease the synthesis of the resultant unfused proteins. In recent years, increasingly more evidence indicates that the N-terminal coding sequence of a procaryotic gene may have a profound effect on translation initiation (4,10–13). Tessier et al (4) have shown that the N-terminal coding sequence of the *c*II protein is involved in the formation of a specific RNA secondary structure that is probably required for maximum efficiency of the *c*II translation initiation region (TIR). It is possible that this structure is somehow important for making the *c*II ribosome binding site highly accessible to the *E.coli* 30S ribosomal subunit (13). Therefore, removal of the *c*II leader coding sequence may prevent efficient translation initiation. One possible solution to this apparent N-terminal coding sequence dependence problem of the *c*II TIR is to replace it with a synthetic TIR. A synthetic TIR could be designed to have minimum potential of secondary structure formation while maintaining the potential to interact effectively with the translation initiation apparatus. Therefore, efficiency of translation initiation may be less dependent on information contained within the N-terminal coding sequence of a target gene. The efficiency of a synthetic TIR could be optimized by designing a theoretically ideal Shine-Dalgarno (SD) sequence and a spacer region (sequence between the SD and the initiator codon) with a length and base composition consistent with known principles governing efficient translation initiation (see 13,14 for reviews). Thus, vectors using synthetic TIRs may be more generally applicable and allow adequate levels of expression of most coding sequences. We describe here the construction and use of a set of vectors containing the λ P_L promoter, synthetic TIRs, and convenient cloning sites. By taking advantage of the power of the polymerase chain reaction (PCR), any target DNA sequence can be rapidly amplified with terminal restriction sites appropriate for its insertion into the vectors (15,16). Single-step directional cloning and high level expression of the target coding sequence should therefore be easily achieved, thereby greatly facilitating the production and purification of unfused gene products.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Nomenclature Used

The translation initiation region (TIR) includes all the information in *cis* that is required for translation initiation (13). It may consist of sequences both upstream and downstream of the the AUG initiator codon (13,17). Thus, one can differentiate upstream and downstream TIR elements (UTIR and DTIR). In this report, synthetic TIR indicates a TIR with a synthetic UTIR. TIR efficiency is the ability of a TIR to initiate translation.

Plasmids and Bacterial Strains

Plasmid pJL6 is an expression vector containing the λ P_L promoter, the phage λ *c*II TIR, and the sequence encoding the N-terminal 14 amino acids of the *c*II protein (5). Plasmids pJL6C, pXC15, pXC16, pXC161, pXC17, pXC171, pXC23, pXC24, pXC24b, pXC241, pXC32, pXC33, pXC33b, pXC34, pXC35, pXC36a, pXC36b, pXC37, pXC46 and pXC47 are expression vectors described in this study. *E.coli* TAP56 (*leu*$^-$, *bio*$^-$, *thi*$^-$, *galK(am)*, Δ*lacU*169, *rpsL*, *supE*44, λ def [Δ*BamHI*, N^+, *c*I857, Δ*cro-bio*]) is a strain that expresses the phage λ *N* transcription antitermination protein and contains the temperature-sensitive phage λ *c*I857 repressor allele. *E.coli* TAP106 is isogenic with TAP56. It contains an inactivating insertion of a Kan^r cassette in the prophage *N* gene (Patterson and Court, in preparation).

Enzymes and Oligonucleotides

Restriction enzyme *Swa*I was from Boehringer Mannheim Corporation (Indianapolis, IN). All other restriction enzymes, the Klenow fragment of *E.coli* DNA polymerase I (*Pol*IK) and T4 polynucleotide kinase were purchased from New England Biolabs (Beverly, MA). T4 DNA ligase and AMV reverse transcriptase were from BRL (Gaithersburg, MD). Taq DNA polymerase was purchased from Cetus Corporation (Emeryville, CA). HPLC-purified oligonucleotides were purchased from Synthecell (Rockville, MD). The PCR, sequencing, and reverse transcription (RT) primers for the amplification of 14-3-3 η chain cDNA (18) are as follows;

*Cla*I.1433	5′-CTCATCGATGGGGGACCGCGAGCAGC-3′
*Nsi*I/*Nde*I.1433	5′-CTCATGCATATGGGGGACCGCGAGCAGC-3′
*Pac*I.1433	5′-CTCATTAATTAATGGGGGACCGCGAGCAGC-3′
*Hind*III.1433	5′-CTCAAGCTTAGTTGCCTTCTCCGGCTT-3′
RT.1433	5′-GGATACTTAGTGAGTGTGG-3′
*Bst*XI.PCR	5′-CTATAAACGCTGATGGAAGCG-3′
*Bst*XI.SEQ	5′-CACATTCCAGCCCTG-3′
*Hind*III.PCR	5′-GCCTATGCCTACAGCATCCAG-3′
*Hind*III.SEQ	5′-GATTTCATACACGGTGC-3′

Underlined regions are complementary to the 14-3-3 η chain cDNA. Primers *Bst*XI.PCR, *Hind*III.PCR, *Bst*XI.SEQ and *Hind*III.SEQ are universal PCR and sequencing primers designed to aid in the screening and analysis of recombinant plasmids. Sequences of oligo cassettes used for the construction of pXC24b, pXC16, pXC161, and pXC241 are given below;

pXC24b.T 5′TAAGGAAGTACTTACA 3′
pXC24b.B 3′**ACGT**ATTCCTTCATGAATGT**AT** 5′

pXC16.T 5′TTTATTTGCATACATTCAATCAATTGTTATCTTAA-GGAAGTACTTACAT 3′
pXC16.B 3′**ACGT**AAATAAACGTATGTAAGTTAGTTAACAATAGA-ATTCCTTCATGAATGTA**GC** 5′

pXC161.T 5′TTTATTTGCATACATTCAATCAATTGTTATCTTAA-CTTTATAAGGAAGTACTTACAT 3′
pXC161.B 3′**ACGT**AAATAAACGTATGTAAGTTAGTTAACAATAGA-ATTGAAATATTCCTTCATGAATGTA**GC** 5′

pXC241.T 5′TTTATTTGCATACATTCAATCAATTGTTATCTTAA-CTTTATAAGGAAGTACTTACA 3′
pXC241.B 3′**ACGT**AAATAAACGTATGTAAGTTAGTTAACAATAGA-ATTGAAATATTCCTTCATGAATGT**AT**-5′

The cohesive termini of the above oligo cassettes are in bold.

General DNA Techniques

Except where specified, DNA manipulations were performed as described in Sambrook et al. (16).

Oligo Replacement

All of the new vectors were constructed using the oligo replacement method described below unless otherwise indicated. This method requires the preparation of oligo cassettes. To prepare oligo cassettes, equal amounts of complementary oligos in TE buffer (pH 7.5) were mixed at a concentration of 1 mg/ml, heated at 90°C for 3 minutes, and incubated at room temperature for 30 minutes. Annealed oligos were diluted to 0.01 mg/ml in TE buffer, phosphorylated in BRL ligation buffer (50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% polyethylene glycol-8000) at 37°C for 30 minutes and then stored at 4°C. Vectors were cleaved with appropriate enzymes and purified from low-melting point agarose gels. One nanogram of oligo cassette was ligated with a fixed amount of vector DNA such that the molar ratio of the oligo to the vector was 2:1. The ligations were done at 14°C overnight. DNA sequencing was performed using Sequenase® according to the manufacturer's recommendation (USB, Cleveland, OH).

PCR Amplifications

Bovine brain poly(A)$^+$ RNA (Clontech, Palo Alto, CA) was reverse transcribed using the RT.1433 reverse transcription primer as described previously (15). The cDNA product from 1 μg of poly(A)$^+$ RNA was diluted to 500 μl with TE buffer (pH 7.5), boiled for 5 minutes and stored at 4°C. Five microliters of the diluted cDNA was used for each 100 μl amplification reaction. In designing PCR primers, an extra 3−5 nucleotides were added to the 5′ ends of the engineered restriction sites to allow more efficient restriction enzyme cleavage (15). Each of the primers *Cla*I.1433, *Nsi*I/*Nde*I.1433, and *Pac*I.1433 is complementary to the N-terminal coding sequence of the 14-3-3 protein and were used in combination with primer *Hind*III.1433 which is complementary to the 14-3-3 C-terminal coding sequence to amplify the 14-3-3 cDNA using conditions recommended by the Cetus Corporation (Emeryville, CA). Where necessary, the magnesium concentration was varied to achieve maximum amplification efficiency. The typical PCR profile for amplification from prepared bovine brain cDNA was: 94°C, (1′) , 60°C, (1′30″), 72°C, (2′). Thirty-five cycles were performed followed by incubation at 72°C for 7 minutes. PCR products were separated on low-melting point agarose gels, purified, cleaved with appropriate restriction enzymes, and gel purified prior to ligation with the vectors. Ligation mixtures were used to transform *E.coli* TAP56 competent cells. Transformants carrying the desired recombinant plasmids were identified with the PCR screening method described below.

PCR screening of bacterial colonies

A single bacterial colony was scraped into 50 μl of 1% Triton X-100, 20 mM Tris-HCl (pH 8.5), 2 mM EDTA and incubated at 95°C for for 10 minutes (15). The bacterial debris was pelleted in a microfuge at 15,000 rpm for 10 minutes. Two microliters of the supernatant were used as template for each 50 μl PCR reaction. Normally, thirty-five cycles of amplification were performed using buffer and reaction conditions recommended by the Cetus Corporation (Emeryville, CA). One unit of Taq DNA polymerase was used for each 50 μl reaction. A typical PCR profile was: 94°C, (1'), 55°C, (1'30''), 72°C, (2'). Ten microliters of each PCR product were analyzed by electrophresis on a 1% agarose gel.

Induction and Protein Analysis

E.coli strains to be induced were grown at 32°C in Luria Broth (LB) (16) containing 100 μg/ml of ampicillin until an OD_{600} of 0.5 and then transferred to 42°C for 4 hours. Aliquots were taken at 0 time and 4 hours after induction. The cell suspensions were adjusted to an OD_{600} of 1.0 with LB and the cell pellet from 1 ml of each suspension was resuspended in 100 μl of SDS loading buffer (16), heated at 100°C for 10 minutes and centrifuged at 15,000 rpm for 10 minutes in a microfuge. Ten microliters of the supernatant was analyzed by polyacrylamide gel electrophoresis under denaturing conditions (SDS-PAGE) (16). Proteins were visualized by Coomassie blue staining. To determine the state of the expressed protein (soluble or insoluble), 10 ml of induced cells were pelleted and resuspended in 1 ml of 50 mM Tris-HCl (pH 7.5). The suspension was sonicated three times at 40 watts for 30 seconds with 1 minute intervals using an ultrasonic microsonicator followed by centrifugation in a microfuge at 4°C at 15,000 rpm for 20 minutes. The supernatant and pellet were separated and analyzed by SDS-PAGE. The relative expression levels of the 14-3-3 protein were estimated using a LKB laser densitometer.

Protein microsequencing

Protein microsequencing was performed as described previously (19). Briefly, proteins were separated by SDS-PAGE (14) and electroblotted to Problott™ membranes (Applied Biosystems) using 10 mM 3-cyclohexylamino-1-propanesulfonic acid (pH 11) in 10% methanol as the transfer buffer. Proteins transferred to the membrane were visualized by Coomassie staining (0.1% Coomassie blue R-250 in 40% methanol/1% acetic acid) and destained in 50% methanol. The bands of interest were then excised and sequenced using an automated sequenator.

RESULTS

I. Expression of 14-3-3 protein using vectors containing the λ *cII* UTIR: Effect of the *cII* leader on expression

Expression of 14-3-3 Protein as a cII Fusion. As part of our investigation into the possible role of the 14-3-3 protein family (18) in the regulation of intracellular phospholipase A_2 (Cheng et al, in preparation), we expressed the cDNA encoding the bovine brain 14-3-3 protein in *E.coli*. This enabled us to prepare highly purified protein for biochemical analysis. A large number of proteins have been successfully overproduced using the pJL6 vector system (1). They are usually expressed as fusions to the *cII* leader, using the highly efficient λ *cII* TIR to direct translation

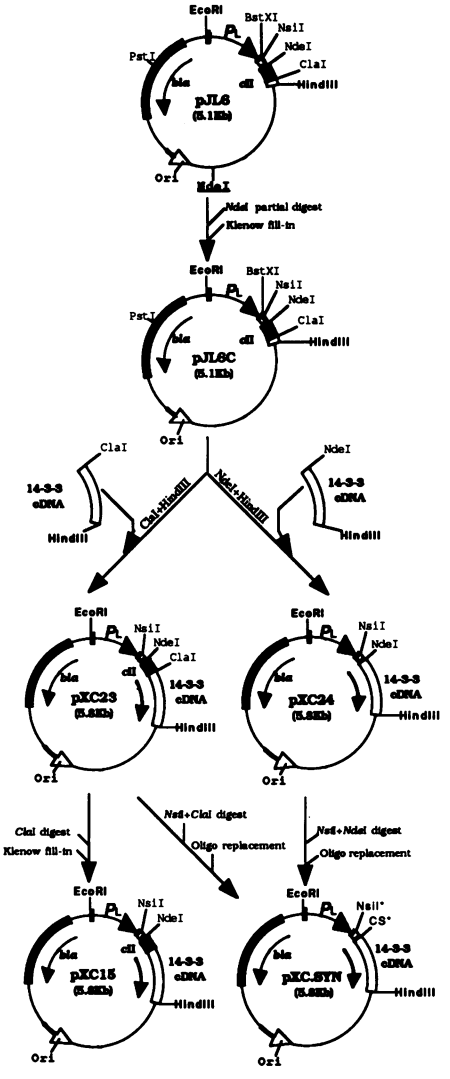

Figure 1. Construction of P_L promoter expression vectors. Maps are not drawn to scale. Important unique restriction sites are marked. Plasmid pJL6 (5) DNA has two *NdeI* sites. The underlined *NdeI* site was eliminated by *NdeI* partial digestion of pJL6 DNA followed by *E.coli* DNA polymerase I Klenow fragment (*pol*IK) treatment and ligation. A plasmid retaining the desired *NdeI* site was selected and named pJL6C. Plasmid pJL6C DNA was digested with *ClaI*+*Hin*dIII and *NdeI*+*Hin*dIII respectively and ligated with PCR amplified 14-3-3 cDNA predigested with the same set of restriction enzymes. The resultant plasmids were named pXC23 and pXC24, respectively. In pXC23, the 14-3-3 cDNA is fused out of frame with the λ *cII* protein N-terminal coding sequence. To make 14-3-3 cDNA in frame with the *cII* leader coding sequence, pXC23 DNA was cleaved with *ClaI*, treated with *pol*IK in the presence of all four dNTPs and religated to generate plasmid pXC15. In pXC24, the *cII* N-terminal coding sequence was removed and therefore 14-3-3 cDNA is directly fused to the *cII* TIR through the *NdeI* site. pXC23 and pXC24 were then used as parental plasmids to construct pXC.SYN which is a general name for all the vectors contructed in this study using the oligo replacement method. To construct pXC16, pXC161, pXC17 and pXC171, plasmid pXC23 was digested with *NsiI* and *ClaI*, gel purified, and ligated with the corresponding oligo cassettes shown (Material and Methods or Fig.3.). To construct of all other vectors, plasmid pXC24 was digested with *NsiI* and *NdeI*, gel purified and ligated with the corresponding oligo cassettes shown (Material and Methods or Fig.3.). P_L, leftward major promoter of phage λ; *bla*, β-lactamase gene; Ori, origin of DNA replication; Kb, kilobase pair(s); *cII*, N-terminal coding sequence for phage λ *cII* proein. *NsiI**: the *NsiI* recognition sequence at this position is eliminated in pXC36, pXC37, pXC46 and pXC47, but a second *NsiI* site is available to be used as the 5' cloning site in these vectors; CS* represents the 5' cloning sites. (see Fig.4 and Table 1 for details). The 14-3-3 cDNA sequence was determined and shown to be identical to the published sequence (18).

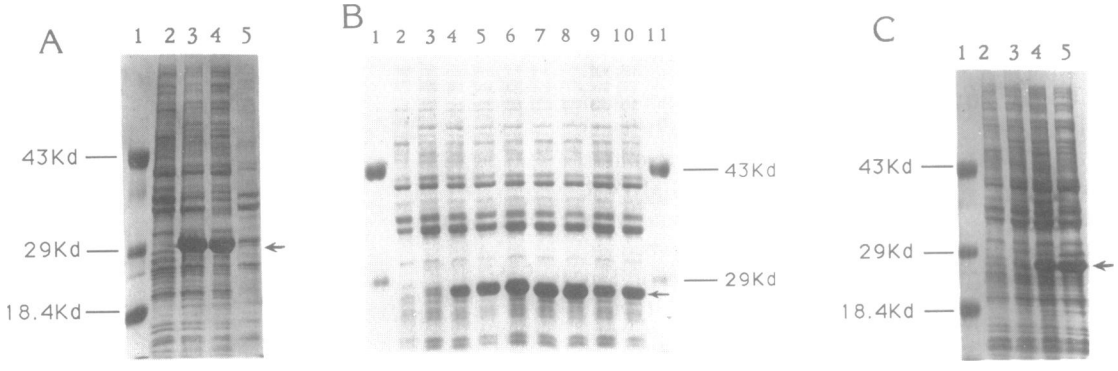

Figure 2. (**A**) SDS-PAGE (10−20%) analysis of 14-3-3 protein expressed as a fusion with the λ *c*II N-terminal 14 amino acids. *E.coli* TAP56/pXC15 was induced and analyzed as described in Materials and Methods. The *c*II::14-3-3 fusion protein comigrates with the 29Kd molecular weight marker and is indicated by an arrow. Lane: 1, molecular weight markers; 2, total cell lysate (TCL) of uninduced cells; 3, TCL of induced cells; 3, supernatant of TCL; 4, pellet of TCL. (**B**) SDS-PAGE (12.5%) analysis of 14-3-3 protein expressed using pXC24 and the synthetic TIR vectors. Lanes: 1,11: molecular weight markers; 2, TCL of uninduced *E.coli* TAP56/pXC24; 3−10, TCL of induced *E.coli* TAP56 containing pXC24, pXC36b, pXC36a, pXC37, pXC34, pXC35, pXC32, and pXC33, respectively. The 14-3-3 protein migrates as 28Kd and is indicated by an arrow. (**C**) SDS-PAGE (10−20%) analysis of 14-3-3 protein expressed using pXC46 and pXC47. Lane: 1, molecular weight markers; 2, TCL of uninduced *E.coli* TAP56/pXC24; 3, TCL of induced *E.coli* TAP56/pXC24; 4, TCL of induced *E.coli* TAP56/pXC46; 5, TCL of induced *E.coli* TAP56/pXC47; The 14-3-3 protein migrates as 28Kd and is indicated by an arrow.

Table 1. Modification of the TIR of pXC15 and its effect on the expression of the 14-3-3 protein

Vector	5′ cloning site	t_{R1}	Enhancer	SD sequence^	Spacer sequence@	Expression (%TCP)*
PXC15	None	wt	−	*c*II	UACUUACAU	>50**
PXC24	*Nde*I	wt	−	*c*II	UACUUACAU	1.5
PXC24b	*Nde*I	Δ	−	*c*II	UACUUACAU	1.5
PXC241	*Nde*I	wt	+	*c*II	UACUUACAU	1.5
PXC16	*Cla*I	wt	−	*c*II	UACUUACAUCG	N.D.
PXC161	*Cla*I	wt	+	*c*II	UACUUACAUCG	N.D.
PXC17	*Cla*I	Δ	−	Syn	AAAAACAUCG	1.5
PXC171	*Cla*I	Δ	+	Syn	AAAAACAUCG	1.5
PXC32	*Pac*I	Δ	−	Syn	UUAAUUAAU	15
PXC33	*Pac*I	Δ	+	Syn	UUAAUUAAU	18
PXC33b	*Pac*I	Δ	+	Syn	UUAAUUA	18
PXC34	*Nde*I	Δ	−	Syn	AAAAACAU	28
PXC35	*Nde*I	Δ	+	Syn	AAAAACAU	40
PXC36a	*Nsi*I/*Nde*I	Δ	−	Syn	AAAAAA	32#
PXC36b	None	Δ	−	Syn	AAAAAAAU	20
PXC37	*Nsi*I/*Nde*I	Δ	+	Syn	AAAAAA	48#
PXC46	*Swa*I/*Nsi*I	Δ	−	Syn	AUUUAA	25#
PXC47	*Swa*I/*Nsi*I	Δ	+	Syn	AUUUAA	33#

N.D., not detectable; t_{R1}, a phage λ transcription terminator; wt, wild type; Δ, partially deleted. Enhancer, UUAACUUUA; ^, the sequences of *c*II and Syn SD are AGGAAG and AGGAGG, respectively; @, spacer sequence is followed immediately by the AUG initiator codon; %TCP*, percentage of total cell protein; **, protein expressed as a *c*II fusion; #, protein expressed as a Met-His fusion;

initiation. We first expressed the bovine brain 14-3-3 η chain in this vector. The construction of plasmid pXC15, from which 14-3-3 protein is expressed as a *c*II fusion protein, is described in Fig.1. As shown in Fig.2A (lane 3), a protein of the size expected for the *c*II::14-3-3 fusion protein accumulated to more than 50% of the total cell protein (TCP) after 4 hours of induction of strain *E.coli* TAP56/pXC15. The identity of this induced protein was comfirmed by protein microsequencing of its N-terminus (Materials and Methods). It was found that more than 90% of the expressed *c*II fusion protein was soluble (Fig.2A, lanes 4 and 5).

Expression of 14-3-3 as an Unfused Protein. Since the presence of the *c*II leader may affect the structure and/or activity of a target protein, we next removed the *c*II leader in an attempt to express unfused 14-3-3 protein. Two plasmids (pXC24 and pXC16) were constructed for this experiment. As shown in Fig.1, plasmid pXC24 was constructed by insertion of the 14-3-3 cDNA between the *Nde*I and *Hind*III sites of vector pJL6C. Plasmid pXC16 was constructed by replacing the sequence between the *Nsi*I and *Cla*I sites of pXC23 (Fig.1) with the oligo cassette shown in Materials and Methods. Plasmid pXC16 differs from pXC24 by two nucleotides (CG) which are located immediately upstream of the AUG initiator codon of the 14-3-3 mRNA transcribed from pXC16 and which are absent from the 14-3-3 mRNA transcribed from pXC24 (Table 1). Both pXC24 and pXC16 are expected to express the 14-3-3 protein in its unfused form under the influence of the modified λ *c*II TIR. *E.coli* TAP56/pXC24 and *E.coli* TAP56/pXC16 were induced and analyzed as described in Materials and Methods. The expressed 14-3-3 protein was detectable on a Coomassie stained SDS-PAGE gel using pXC24 (Fig2B, lane 3), but at a much reduced level compared with the *c*II fusion, indicating that removal of the *c*II leader caused a dramatic decrease in expression of the resultant unfused 14-3-3

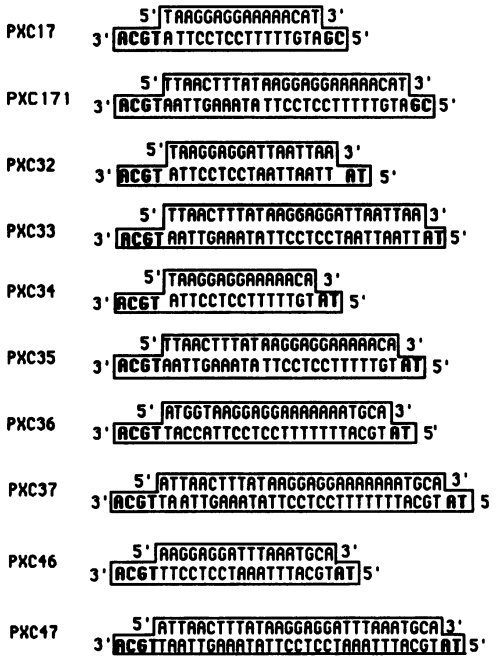

Figure 3. Oligo cassettes used for the construction of synthetic TIR vectors. The sequence of each oligo and the cohesive termini of each cassette are presented. Each oligo cassette possesses either the *NsiI* and *NdeI* or *NsiI* and *ClaI* cohesive termini (in bold) which are compatible with the (*NsiI*+*NdeI*) or (*NsiI*+*ClaI*) digested vectors. Each oligo cassette contains sequence information corresponding to a SD sequence, a spacer region between the SD and the AUG start codon, and one or two 5′ cloning sites.

protein. No 14-3-3 protein was detectable on a Coomassie stained SDS-PAGE gel using pXC16 (data not shown), suggesting that the CG dinucleotide in the pXC16 TIR causes a further decrease in the efficiency of the modified *c*II TIR in pXC24.

It has been reported that a sequence (5′-UUAACUUUA-3′) from phage T7 mRNA can act as an enhancer of translation of the *lacZ* gene in *E.coli* (20). To test whether this sequence could increase the expression of unfused 14-3-3 protein using pXC24 and pXC16, plasmids pXC241 and pXC161 were constructed (oligo cassettes used are shown in Materials and Methods). In these plasmids, the enhancer sequence was inserted immediately upstream of the *c*II SD sequence. No significant change in 14-3-3 protein expression was observed (data not shown, see Table 1).

II. Design and construction of vectors containing a synthetic TIR

Necessity and Design. Our data show that the *c*II TIR-based vectors can express 14-3-3 protein as a *c*II fusion to a very high level but removal of the *c*II leader leads to a dramatic decrease in the expression of the resultant unfused 14-3-3 protein. One possible explanation for this result is that the *c*II TIR requires information contained within the *c*II leader for optimum efficiency (4). To circumvent such a requirement, we designed synthetic TIRs with a consensus SD sequence. Protein expression levels from vectors containing such synthetic TIRs might be expected to be more independent of the N-terminal coding sequence of target genes or cDNAs. The vectors could therefore be used for more efficient expression of unfused 14-3-3 protein as well as for the expression of other gene products.

We designed a series of vectors each having a different restriction site as the 5′ cloning site. The term 5′ cloning site

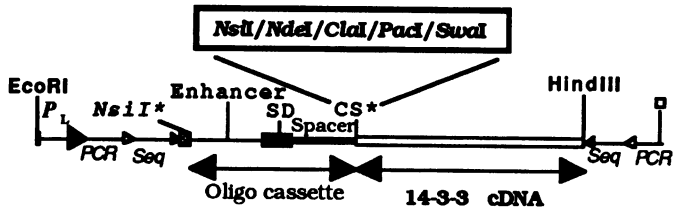

Figure 4. Diagram of the promoter and TIR regions of the synthetic TIR vectors. Map is not drawn to scale. The region between restriction sites *EcoRI* and *HindIII* which includes the promoter and TIR region of the synthetic TIR vectors are illustrated in more detail. CS*: represents the 5′ cloning sites. Only one or two of the five restriction sites (boxed) are present in each vector (see Table 1 for details). *NsiI**: the *NsiI* recognition sequence at this position is eliminated in pXC36, pXC37, pXC46 and pXC47, but a second *NsiI* site is available to be used as the 5′ cloning site in these vectors. The universal PCR-screening primers (*BstXI*.PCR and *HindIII*.PCR) and sequencing primers (*BstXI*.SEQ and *HindIII*.SEQ) are designated *PCR* and *Seq* and are indicated by open and solid arrows, respectively.

here is defined as a restriction site on the vector that provides at least part of an AUG initiator codon and is used to fuse the 5′ end of a target DNA sequence to the vector. This design strategy requires that part of the 5′ cloning site recognition sequence be present in the spacer region of the synthetic TIR. Since it is known that the sequence of the TIR spacer can influence the efficiency of translation initiation (14), we tested several different restriction sites for suitability. The criteria used for the selection of the 5′ cloning sites are: (a) frequency of appearance of the site in natural DNAs; (b) base composition of the site, and (c) potential to allow expression of unfused proteins. It is generally accepted that the spacer sequence between the SD and the AUG initiator codon should be AT-rich and between 4 and 12 bp in length (14,21). Restriction sites *Pac*I(TTAAT'TAA), *Nde*I(CA'TATG), *Nsi*I(ATGCA'T), *Swa*I(ATTT'AAAT) and *Cla*I(AT'CGAT), all of which can provide at least part of a translation initiation codon, were selected and tested. The *Pac*I and *Swa*I sites are unique in two ways. First, they are among the few commercially available restriction enzymes with a recognition sequence of eight nucleotides. Therefore, the chance of an internal *Pac*I or *Swa*I site within target DNA sequences will be rare. Secondly, the *Pac*I and *Swa*I recognition sequences will generate spacers between the SD sequence and the AUG initiator codon which fall within the theoretically ideal range of length and base composition (21). We also designed vectors with and without the translational enhancer sequence from the phage T7 gene 10 TIR in order to evaluate the effect of this element on expression of proteins using the *P*L promoter-based vectors (20). The vectors described below are referred to either by their pXC number or by their 5′ cloning site. For example, pXC32 and pXC33 have *Pac*I as their 5′ cloning site and are referred to as *Pac*I vectors. Similarly, pXC46 and pXC47 utilize *Swa*I as their 5′ cloning site and are therefore called *Swa*I vectors.

Construction and use of Vectors with a Synthetic TIR. All plasmids were constructed by replacing the UTIR of the parental plasmids with oligo cassettes. Plasmid pXC23 was used to construct pXC17 and pXC171. pXC24 was used as the parental plasmid for the construction of pXC32 through pXC37, pXC46, and pXC47. Since pXC23 and pXC24 contain the 14-3-3 cDNA, each of the new vectors is expected to express 14-3-3 protein either in its unfused form or, in the case of pXC36, pXC37,

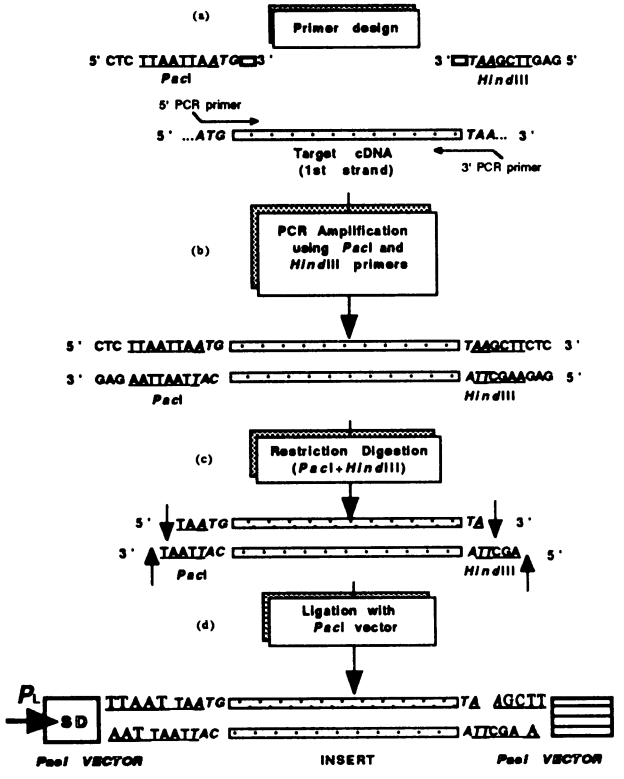

Figure 5. Use of synthetic TIR vectors. The use of vectors described in this study is illustrated using a *Pac*I vector as an example. The sequences of the 5′ and 3′ cloning sites are underlined. The start and stop triplets of the target cDNA coding sequence are indicated by italics. Vector sequences are outlined. Generally, four steps are needed to insert a target into the desired vector. They are: (a) Primer design: PCR primers are usually 20 to 30 bp in length with 15 to 18bp complementary to 5′ or 3′ end of the target coding sequence. The recognition sequence of a desired restriction site is added at the 5′ end of the primer with an extra 3 to 5 bp that are 5′ to the restriction site to improve cleavage effciency (15). An AUG start codon should be included if it is not present in the target sequence. The AUG should be completely or partially contained within the 5′ cloning site recognition sequence so that proper spacing between the SD sequence and the initiator codon is maintained. (b) cDNA amplification: cDNA template is prepared as described in Materials and Methods and amplified using the designed primers. (c) Restriction digestion: PCR products are purified and digested with appropriate enzymes followed by phenol extraction and ethanol precipitaion. (d) Ligation with vector: cleaved PCR insert is ligated with the vector predigested with the same set of enzymes.

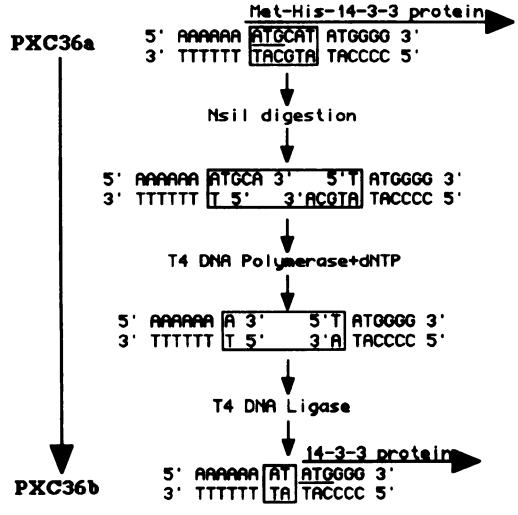

Figure 6. Conversion of plasmid pXC36a to pXC36b. Plasmid pXC36a, which will express 14-3-3 protein as a Met-His fusion, was digested with *Nsi*I followed by T4 DNA polymerase treatment in the presence of all four dNTPs to flush the ends. T4 DNA ligase was used to ligate the resultant blunt ends to generate pXC36b. The evolution of the *Nsi*I site (ATGCAT) in pXC36a after each step can be seen by following the boxed sequence. The underlined ATG in pXC36a and pXC36b corresponds to the AUG start codon of 14-3-3 mRNA.

pXC46, and pXC47, as Met-His fusions (Table 1). Details of the construction of the vectors are given in Fig.1. Oligo cassettes used are shown in Fig.3. The promoter and TIR region of the synthetic TIR vectors are shown in more detail in Fig.4. The nucleotide sequence of the TIR of each vector was verified by DNA sequence analysis. The 5′ cloning sites of the vectors are given in Table 1, as are other important features of their TIRs. As can be seen in Table 1, the odd numbered vectors in the series pXC32 through pXC37 and pXC47 contain a translational enhancer sequence derived from phage T7 gene 10 mRNA (20); all others lack this sequence.

The use of the vectors is illustrated in detail in Fig.5 using the *Pac*I vector as an example. In general, target DNA sequences to be expressed are PCR-amplified using customized primers designed to adapt the sequence to the specific vector to be used. After cleavage of the PCR product with the appropriate enzymes, it can be directionally inserted into the vector. All that is required is knowledge of short segments of sequence at the 5′ and 3′ ends of the target. It should be noted that proteins expressed using the *Nsi*I vectors will have two extra N-terminal amino acids (Met-His) unless a second in vitro manipulation is performed. This is illustrated in Fig.6 for the 14-3-3 coding sequence. It should also be noted that although the *Swa*I vectors (pXC46 and pXC47) were used to express the 14-3-3 protein as a Met-His fusion in this study, both vectors can be used to express unfused gene products to equivalent levels without requiring a second in vitro manipulation (data not shown). Therefore, all vectors except pXC36 and pXC37 can be used to produce proteins with authentic N-termini in a single step.

The ability to design universal PCR and sequencing primers for the vectors facilitates the identification and analysis of recombinants. Using the universal PCR primers (*PCR*) (Fig.4), transformant colonies can easily be screened for the presence of inserts (Materials and Methods). Transformants containing a vector with the desired insert will yield a PCR product that is 280 bp larger than the insert while those containing a vector without an insert will not give any products larger than 280 bp (the distance between the two *PCR* primers on the vectors). Since the universal sequencing primers (*Seq*) are nested to their corresponding *PCR* primers, it is also possible to sequence the PCR-amplified insert directly using the *Seq* primers (Fig.4).

Expression of the 14-3-3 Protein in the Newly Constructed Vectors. The results of expression experiments with the vectors containing synthetic TIRs are shown in Fig.2B and Fig.2C. The identity of the expressed proteins was confirmed by N-terminal amino acid sequence analysis. A summary of the estimated expression levels of the 14-3-3 protein using each expression vector is included in Table 1. As can be seen, modification of the TIR of the original expression vector (pXC24) has resulted in a significant increase in expression of unfused 14-3-3 protein. Although most of the new vectors containing a synthetic TIR give high level expression of unfused 14-3-3 protein (15−48% TCP),

there are some interesting differences in expression levels among the vectors. First, as had been seen with the the *c*II TIR vectors, the presence of the dinucleotide CG immediately upstream of the AUG start codon of 14-3-3 mRNAs transcribed from pXC17 and pXC171 has a profound negative effect on the expression of 14-3-3 protein. When the CG was removed, expression levels increased from 1.5% TCP to 40% TCP (pXC171 vs pXC35) (Table1). Other changes in the spacer regions of the vectors do not affect the expression of 14-3-3 protein in such a dramatic manner. The CG effect could result from the altered spacer base composition, increased spacer length, or both. Secondly, the presence of a translational enhancer sequence from phage T7 increases the expression of 14-3-3 protein though not significantly. We observed no decreases in expression attributable to the addition of the enhancer sequence (Table 1).

In addition to differences in the SD and spacer regions, one potentially significant difference between the original vector (pXC24) and the synthetic TIR vectors is that in the latter, part of the phage λ t_{R1} transcription terminator (22) is deleted. It is not known whether this deletion will affect t_{R1}-mediated transcription termination. However, two lines of evidence indicate that the differences in expression levels between pXC24 and the synthetic TIR vectors are not attributable to this partial deletion of the t_{R1} terminator. First of all, expression of the 14-3-3 protein from pXC24 was the same in *E.coli* strain TAP56, which produces the phage λ *N* antitermination protein, and in the λ N^- *E.coli* strain TAP106 (data not shown). The *N*-mediated antitermination system efficiently suppresses the t_{R1} transcription terminator (22). This suggests that transcription termination is not a factor in the expression of the 14-3-3 protein. Secondly, a pXC24 derivative (pXC24b) which carries the same partial deletion of the t_{R1} terminator gave the same level of expression of 14-3-3 protein as pXC24, indicating that the partial deletion of t_{R1} does not improve the expression of 14-3-3 protein when the *c*II TIR is present (Table 1).

DISCUSSION

Translation initiation often acts to limit the overall efficiency of gene expression in *E.coli* but has proved difficult to optimize in a generally applicable manner because the details of the process are not completely understood. High level expression has commonly been achieved by fusing target genes to strong natural TIRs such as that of the phage l *c*II gene. However, increasingly more evidence suggests that the efficiency of a TIR may depend strongly on information contained within the N-terminal portion of the coding sequence (4,10−13). Therefore, poor translation initiation may result when attempting to express unfused heterologous proteins using a natural TIR. Our results with 14-3-3 protein expression using the *c*II TIR support this hypothesis. In designing the vectors described in this report, we attempted to both idealize the TIRs and incorporate maximum versatility. The vectors utilize *Pac*I, *Nde*I, *Cla*I , *Nsi*I and *Swa*I restriction sites to provide all or part of an AUG initiator codon and facilitate the precise fusion of target DNA sequences to vector transcriptional and translational signals. In addition, universal sequencing and PCR primers have been designed to aid in colony PCR screening and DNA sequence analysis of recombinant plasmids.

We have excluded the possibility that increased expression levels of 14-3-3 protein from the synthetic TIR vectors are due to a t_{R1} deletion. Therefore, we think that the observed differences in expression of the 14-3-3 protein between the two types of vectors (*c*II TIR vector and synthetic TIR vector) are due to variations in the SD and spacer regions which consequently affect the translation initiation efficiency, the functional half-life of the mRNA, or both (14,23−25). The precise mechanism of how the *c*II leader coding sequence affects expression of proteins using the *c*II TIR is not clear. Three hypotheses can be put forward to explain this. The simplest explanation for this is that the *c*II leader coding sequence is part of the *c*II TIR. Therefore, removal of this sequence would directly decrease ribosome binding. Another possibility is that removal of the *c*II leader coding sequence has decreased the accessibility of the resulting TIR to ribosomes. It has been found that in many cases, secondary structure formation involving TIRs inhibits translation initiation (26,27). However, analysis of λ *c*II::α1AT gene fusions has shown that increased expression correlates with the stabilization of a particular computer predicted RNA secondary structure involving the *c*II leader coding sequence (4). Thus, in this particular case, a specific secondary structure may somehow be required to expose the ribosome binding site. When the *c*II leader coding sequence is deleted, a new secondary structure that now inhibits translation initiation may form. The third possibility is that deletion of the *c*II leader coding sequence has decreased the stability of the resulting mRNA. It has been shown that changes near the beginning of the coding sequence of an mRNA can influence the degradation rate of that message (23,25). Further experiments are required to distinguish among these possibilities. The synthetic TIRs in our vectors were designed to minimize the potential of secondary structure formation and thus are less likely to form inhibitory structures when a target coding sequence is fused downstream. In addition, the TIRs were designed to effectively interact with the translation initiation complex without requiring contributions from sequences downstream of the trsanslation initiation codon. The success of our strategy is demonstrated by the significant increase in expression of the unfused 14-3-3 protein when the synthetic TIR vectors were used. Although most of the vectors express the 14-3-3 protein to a high level (>15% total cell protein), none of them yielded the extremely high level expression of 14-3-3 protein seen with the *c*II fusion construct. Thus, in generalizing the TIRs, we probably compromised optimum efficiency. To achieve maximum translation initiation of 14-3-3 protein using the synthetic TIR vectors, alterations in the N-terminal coding sequence may be required (28). However, the expression levels obtained using our vector system are more than adequate for most applications.

The only differences among the synthetic TIR vectors are in the spacer regions, in the presence or absence of two extra N-terminal codons (see *Nsi*I vectors), and in the presence or absence of the T7 gene 10 translational enhancer. Therefore, differences in expression levels must be due to these variations. It had been shown earlier that the presence of C and/or G at the −1 or −2 position (relative to the AUG initiator codon) is unfavorable for maximum translation efficiency (18,26). Therefore, the most plausible explanation for the negative effect of the sequence CG in the spacer region of the TIRs in pXC16, pXC161, pXC17 and pXC171 is that it reduces translation initiation of the 14-3-3 mRNA. We have found that the dinucleotide CG at position −2 and −1 does not always have a negative effect on the expression of a target protein. We have used pXC17 and another vector which utilizes *Cla*I as the 5′ cloning site to express other proteins to a high level (Patterson et al; in preparation). Thus, the negative effect of CG at the −1 and −2 position appears to be gene

dependent. We also tested whether the reported translational enhancer sequence from the phage T7 gene 10 TIR could improve expression of the 14-3-3 protein in vectors utilizing the *c*II UTIR and whether it could further increase the efficiency of the synthetic TIRs. Our results indicate that the translational enhancer sequence has no significant effect (<2 fold) on the expression of the 14-3-3 protein. However, we have observed 5 to 10 fold increases in the expression of human DNA polymerase β and the yeast APN1 protein when the enhancer is present (Patterson et al; unpublished data). Other investigators have shown that this same enhancer sequence can increase the translational efficiency of the *E.coli lacZ* gene by a factor of 110 (20). We conclude that enhancer effects are also gene dependent.

Although the enhancer sequence did not negatively affect expression levels, it is probably most prudent to test vectors with and without this sequence. In fact, the ease with which DNA sequences can be PCR-amplified and inserted into the vectors coupled with the ability to use universal PCR and sequencing primers to aid in identification and verification of recombinant plasmids makes a vector survey approach to protein expression problems both attractive and practical. In addition to the 14-3-3 protein described here, we have successfully overexpressed a number of other eukaryotic proteins using this vector system.

In summary, we have constructed a set of P_L-based expression vectors which permit single-step directional cloning and high level expression of PCR amplified DNA coding sequences. Our results suggest that a synthetic TIR with consensus translational initiation signals is a good alternative to strong natural TIRs for high level expression of unfused gene products using phage λ P_L promoter-based vectors. The vectors we describe here should be widely applicable for heterologous protein expression in *E.coli*.

ACKNOWLEDGEMENTS

We would like to thank Dr. Nat Sternberg and Dr. Lynn Enquist for critically reading the manuscript. We thank Ms. Sylvia Stack for her excellent assistance in sequencing the 14-3-3 cDNA and Mr. Wayne Little for his excellent technical help throughout this work. We are indebted to Ms. Jeanne Corman and Dr. Ram Seetharam for performing the N-terminal amino acid sequence analysis and Dr. Richard Yates for his help in estimating the protein expression levels by laser densitometry.

REFERENCES

1. Samuel, K.P., Ascione, R., Kottaridis, S.D., Seth, A.K., Lautenberger, J.A., Zuber, M., Strouboulis, J., and Papas, T.S. (1990) *Genet. Anal. Tech. Appl.* **7**, 178–208.
2. Chen, S.M., Takiff, H.E., Barber, A.M., Dubois, G.C., Bardwell, J.C.A., and Court, D.L. (1990) *J. Biol. Chem.* **265**, 2888–2895.
3. Wang, H., McConnell, D.J., and O'Mahony, D.J. (1990) *Nucl. Acids Res.* **18**, 1070.
4. Tessier, L.H., Jallat, S., Sauvageot, M., Crystal, R.G., and Courtney, M. (1986) *FEBS letters* **208**, 183–188.
5. Lautenberger, J.A., Court, D., and Papas, T.S. (1983) *Gene* **23**, 75–84.
6. Shimatake, H., and Rosenberg, M. (1981) *Nature* **292**, 128–132.
7. SenGupta, D.N., Zmudzka, B.Z., Kumar, P., Cobianchi, F., Skowronski, J., and Wilson, S.H. (1986) *Biochem. Biophys. Res. Commun.* **136**, 341–347.
8. Popoff, S.C., Spira, A.I., Johnson, A.W., and Demple, B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4193–4197.
9. Cheng, X., Bunville, J., and Patterson, T.A. (1992) *Nucl. Acids Res.* **20**, 370.
10. Tomich, C.S.C., Olson, E.R., Kaytes, P.S., Rockenbach, S.K., and Hatzenbuhler, N.T. (1989) *Nucl. Acids Res.* **17**, 3179–3197.
11. Sprenggart, M.L., Fatscher, H.P., and Fuchs, E. (1990) *Nucl. Acids Res.* **18**, 1719–1723.
12. Petersen, G.B., Stockwell, P.A., and Hill, D.F. (1988) *EMBO J.* **7**, 3957–3962.
13. McCarthy, J.E.G., and Gualerzi, C. (1990) *Trends Genet.* **6**, 78–85.
14. Gold, L. (1988) *Ann. Rev. Biochem.* **57**, 199–233.
15. Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. (1990) PCR Protocols: A Guide to Methods and Applications. Academic press, Inc., San Diego, CA.
16. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
17. Dreyfus, M. (1988) *J. Mol. Biol.* **204**, 79–94.
18. Ichimura, T., Isobe, T., Okuyama, T., Takahashi, N., Araki, K., Kuwano, R., and Takahashi, Y. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7084–7088.
19. Matsudaria, P (1987) *J. Biol. Chem.* **262**, 10035–10038.
20. Olins, P.O., and Rangwala, S.H. (1989) *J. Biol. Chem.* **264**, 16973–16976.
21. Hartz, D., McPheeters, D.S., and Gold, L (1991) *J. Mol. Biol.* **218**, 83–97.
22. Court, D., Brandy, C., Rosenberg, M., Wulff, D.L., Behr, M., Mahoney, M, and Izumi, S. (1980) *J. Mol. Biol.* **138**, 231–254.
23. Guillerez, J., Gazeau, M., and Dreyfus, M. (1991) *Nucl. Acids Res.* **19**, 6743–6750.
24. Emory, S.A. and Belasco, J.G. (1990) *J. Bacteriol.* **172**, 4472–4481.
25. Petersen, C. (1987) *Mol. Gen. Genet.* **209**, 179–187.
26. Gross, G., Mielke, C., Hollatz, I., Blocker, H., and Frank, R. (1990) *J. Biol. Chem.* **265**, 17627–17636.
27. Smit, M.H.D. and Duin, J.V. (1990) In Cohn, W.E. and Moldave, K. (ed.), Progress in Nucleic Acid Research and Molecular Biology. Academic Press, Inc., San Diego, Vol. **38**, pp. 1–35.
28. Bucheler, U.S., Werner, D., and Schirmer, R.H. (1990) *Gene* **96**, 271–276.