Gene, 75 (1989) 323–327
Elsevier

GEN 02872

New plasmid vectors for high level synthesis of eukaryotic fusion proteins in *Escherichia coli*

(Recombinant DNA; plasmid; multiple cloning site; promoter; protein degradation; *Acanthamoeba*; myosin)

David L. Rimm and Thomas D. Pollard

Department of Cell Biology and Anatomy, Johns Hopkins School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205 (U.S.A.)

Received by J.L. Slightom: 14 September 1988
Revised: 15 October 1988
Accepted: 21 October 1988

SUMMARY

Production of eukaryotic proteins in *Escherichia coli* has become rather simple since commercially available bacteriophage and plasmid vector systems allow investigators to select the optimal system for their particular problem. A common question is which system to use to produce the largest quantity of soluble recombinant protein with minimal, if any, bacterial protein fused to it. We have constructed a new set of plasmid vectors that produce large amounts of a fusion proteins that contain less than 25 amino acids of bacterial protein. We started with pATH-1, a plasmid expression vector comprised of the *trpEp* promoter and 37 kDa of the TrpE protein followed by a M13mp13 multiple cloning site for insertion of sequences to be expressed. We deleted the majority of the eukaryotic *trpE* sequence to produce a multiple frame, multiple enzyme cloning site, plasmid expression vector set called pRX. Transformation of *E. coli* CAG-456 (Baker et al., 1984) with this vector with an *Acanthamoeba* myosin tail sequence inserted in the correct frame yields a fusion protein that represents 45% of the total soluble protein. We have produced and purified 100 mg of this *Acanthamoeba* myosin-II fusion protein per liter of cell suspension.

INTRODUCTION

A longtime goal of molecular biologists has been to produce eukaryotic products in bacteria. Many systems have been engineered specifically for this purpose (for review see Harris, 1983, or Denhardt and Colasanti, 1987). Originally, the goal of many of these projects was to produce therapeutically or commercially useful products; however, now these systems are also being exploited for studies of the structural and functional properties of the recombinant proteins (Leinwand, 1988).

Correspondence to: Dr. T. Pollard, Dept. of Cell Biology and Anatomy, Johns Hopkins School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205 (U.S.A.) Tel. (301)955-5672; Fax 301-955-4129.

Abbreviations: aa, amino acid(s); Ap, ampicillin; bp, base pair(s); kb, kilobase(s) or 1000 base pairs; MCS, multiple cloning site; nt, nucleotide(s).

0378-1119/89/$03.50 © 1989 Elsevier Science Publishers B.V. (Biomedical Division)

Expression of functional proteins in bacteria has been difficult in that they lack the eukaryotic subcellular processing systems required for post-translational modification and they produce proteolytic enzymes that degrade foreign proteins. Typically, fusion proteins are easier to produce in large quantities than non-fusion proteins, but the bacterial protein portion often renders the eukaryotic protein inactive or insoluble. Some vectors allow production of recombinant proteins with their own N-terminus but the yield is usually small compared to fusion protein yields (Straus and Gilbert, 1985; Putkey et al., 1985). The difference is presumedly due to degradation of the foreign protein inside the bacterial cell. The bacterial N-terminal amino acids seem to have a stabilizing effect on foreign proteins (Hare et al., 1984; Bachmair et al., 1986). The optimal arrangement may be a recombinant protein with a minimal number of bacterial residues at the N-terminus. A small bacterial extension may protect a recombinant protein from bacterial degradation without interfering with the functions of its eukaryotic portion. For example, Courtney et al. (1984) produced a human α-antitrypsin with 17 N-terminal aa from the CII protein of phage λ and artificial polylinker sequence and found that fusion product retained full function.

We have produced a new set of vectors designed to maximize expression of functional protein while minimizing bacterial degradation. They are driven by a strong bacterial promoter, *trpEp*, and the fusion proteins have as few as 18 TrpE aa at their N-termini. Expression is further maximized by the use of *E. coli* strain CAG-456 (Baker et al., 1984) with this new vector.

EXPERIMENTAL AND DISCUSSION

(a) Construction of the pRX vectors

The pATH-1 expression vector (T.J. Koerner, personal communication; Dieckmann and Tzagoloff 1985) was used as the starting material. pATH-1 was produced by fusing the *trpEp* promoter and the first 969 nt encoding the TrpE protein to the multiple cloning site of M13mp13 (Vieira and Messing, 1982). The resulting plasmid has been used frequently, but yields a fusion protein with 37 kDa of bacterial protein that is often found as an insoluble product (Earnshaw et al., 1987). We cut pATH-1 at an *Nru*I site located 51 nt after the *trpE* translation initiation codon. The blunt ends left by *Nru*I were ligated to 8-mer, 10-mer and 12-mer *Eco*RI linkers purchased from Pharmacia (Piscataway, NJ). The plasmids were then transformed into HB101 and grown to select for insertion of the linkers by restriction digestions. The *Eco*RI digested plasmids showed a band at approx. 2.8 kb representing the vector and a second band at approx. 900 bp representing the majority of the coding region of the *trpE* gene. The ~2.8 kb band was cut out, electro-eluted, and ligated. The resulting plasmids, called pRX-1, -2 and -3, were grown up and screened by restriction mapping for size and then sequenced using the dideoxyribonucleotide method with a synthetic primer (Sanger et al., 1977). Sequencing back through the multiple cloning site and past the *trpE* translation initiation codon showed that all three reading frames could be obtained. The reading frames follow the names where the reading frame is defined as the number of bases after the beginning of the codon at

TABLE I

The number of the reading frame required for expression of each usable restriction enzyme in each pRX vector

Vector[a]	*Eco*RI	*Sac*I	*Sma*I	*Bam*HI	*Xba*I	*Sal*I	*Hin*dIII	*Cla*I
	Frame number for each restriction site[b]							
pRX-1	1	2	1	1	1	1	1	2
pRX-2	2	3	2	2	2	N	N	N
pRX-3	3	1	3	3	3	3	3	1

[a] See Fig. 2.

[b] The number shown is the frame number for insertion of foreign DNA where frame number is defined as the number of nucleotides after the beginning of the codon at which the restriction enzyme cuts. The letter 'N' indicates that the site is not available for cloning for expression since there is a 'TAG' stop codon (part of the *Xba*I restriction site) in the reading frame preceding these sites.

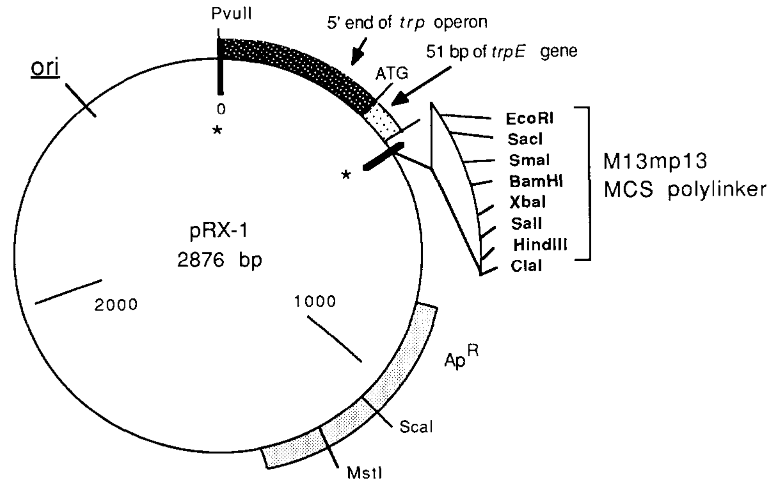

Fig. 1. A restriction map of pRX-1 Restriction sites in the MCS occur only once and are available for cloning. The plasmid is composed of the *trp* operon (dark shading) including 51 bp of coding region (light shading), the M13mp13 MCS polylinker and pBR322. The asterisks denote the borders of the pBR322 derived DNA starting at the *Hin*dIII site of the polylinker and around to the *Pvu*II site. The numbering defines the junction of pBR322 and *trp* operon at 0 and proceeds clockwise. The ampicillin resistance gene (Ap^R) from pBR322 is shaded light grey. Construction of the plasmid is described in section (**a**) in the text.

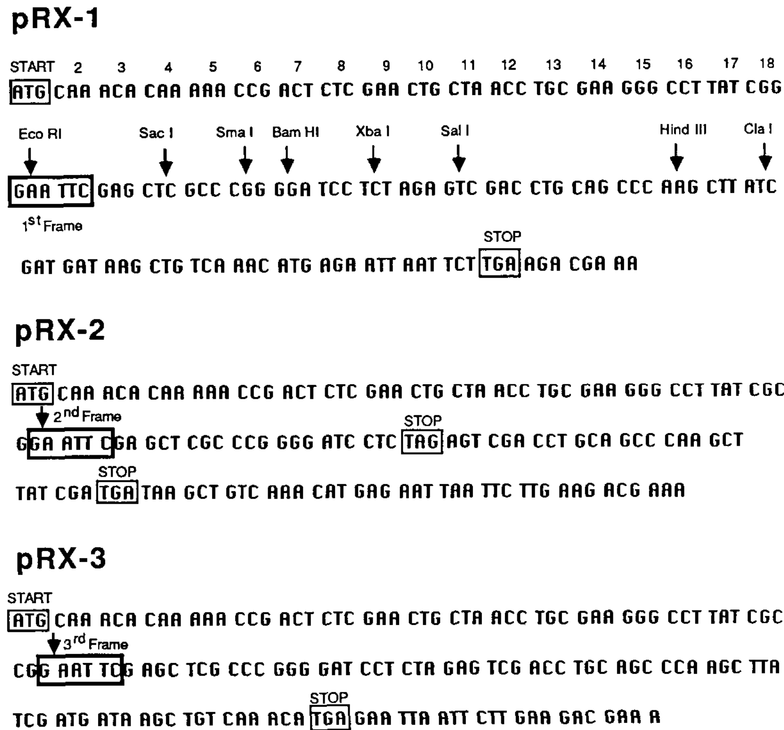

Fig. 2. The DNA sequence corresponding to the N-terminus derived from the *trpE* protein for determination of the N-terminal sequence of derived fusion proteins. The reading frames of all usable enzymes in the MCS, and stop codons of all 3 pRX vectors are shown. The bold box indicates the initial *Eco*RI site and its location in each frame. The other boxes indicate the start and stop sequences in frame in each vector. We have found this diagram useful for determining the exact end sequences of our fusion proteins. Note that a *Pst*I site is present between the *Sal*I site and the *Hin*dIII site but it is not shown since it is not unique. A second *Pst*I site is present in the sequence derived from pBR322 that makes this site unavailable for cDNA cloning.

which the enzyme cleaves (see Table I). Fig. 1 shows a restriction map of pRX-1. Fig. 2 shows the sequence of the *trpE* start and the multiple cloning site polylinker in all three expression frames.

(b) Expression of functional recombinant proteins

The pRX plasmids grow about as well as pBR322 with respect to copy number and, in many cases substantial amounts of fusion protein can be produced using standard host strains such as HB101 or DH-1. We have found that we have consistently been able to produce ten-fold more fusion protein using a stain called CAG-456 [*lac*am *trp*am *pho*am *supC*^{ts} *rpsL*(Sm^R) *mal*am *htpR*165] (Baker et al., 1985). This strain is a mutant at the *htpR* locus which is a regulator of the heat shock response. Baker et al. (1985) showed that cells with this mutation are non-specifically defective in degrading proteins. This mutation is distinct from strains with mutations in the *lon* gene, commonly used in producing fusion proteins. Later work showed a functional *htpR* gene product is required to induce the *lon* gene (Goff and Goldberg, 1985). It is not yet clear if the decreased amount of the *lon* gene product, protease La, is the sole factor responsible for decreased protein degradation; however, *lon*$^-$ strains without the *htpR* mutations typically result in lower fusion protein yields.

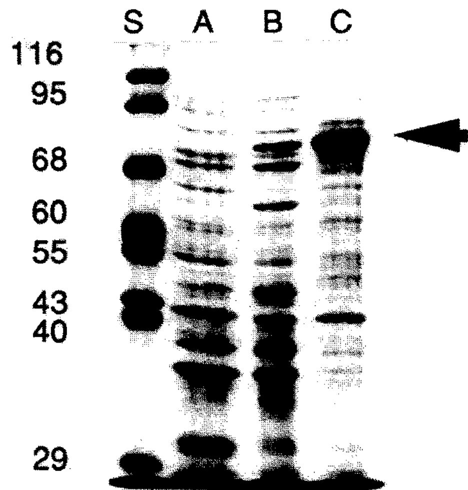

Fig. 3. Gel electrophoresis (Laemmli, 1970) of soluble bacterial extracts stained with Coomassie Blue. Lanes: (S) standard proteins with molecular weights in kDa, (A) pRX-1 in HB101 with no insert, (B) myosin tail clone in pRX-1 in HB101, and (C) *Acanthamoeba* myosin-II tail clone in pRX-1 in CAG-456. The arrow indicates the bands corresponding to the myosin tail fusion protein. The cells are grown, induced and harvested as described in section (**b**) of the text. Gel samples are made by pelleting 1 ml of a cell suspension for 10 s in a microfuge, washing with 1 ml of 10 mM Tris · HCl (pH 7.5), resuspending in 100 μl of Laemmli (1970) sample buffer, and boiling for 3 min before loading 10 μl on a 0.1% SDS–7.5% polyacrylamide gel.

CAG-456 is slow-growing and inefficient for producing plasmid DNA. Therefore, we transform into it to boost expression only after we have verified fusion protein expression in another strain. Transformation is done according to standard $CaCl_2$ methods (Maniatis et al., 1982) except heat shock is to 37 °C if at all and the cells are allowed to grow in the absence of antibiotic for 1.5–2 h instead of the normal 30 min. Preparations of fusion protein are typically initiated by growing the transformed CAG-456 cells to stationary phase at 30 °C in LB broth. The cells are then inoculated into 4 vols. M9-CA minimal medium with 50 μg Ap/ml (Maniatis et al., 1982). The cells are grown for 3 h at 30 °C, then induced with indolyl acrylic acid (Sigma Chemical, St. Louis, MO) to a final concentration of 10 μg/ml. After 5 h more, the cells are harvested and washed in preparation for gel electrophoresis or purification of the fusion protein.

We have used the pRX-1 vector in CAG-456 to produce a myosin tail fusion protein. The fusion product contains 18 aa of TrpE protein, 3 aa of synthetic linker and 568 aa of *Acanthamoeba* myosin-II. It is purified from sonic lysates in a soluble form with a yield of 5 mg of protein per 50 ml of suspension culture of cells in late log phase. The fusion protein represents approx. 45% of the total soluble protein as shown in Fig. 3. After growth and induction as described above, the cells are lysed by sonication in 10 mM Tris · HCl (pH 8), 50 mM EDTA, 8% sucrose, 0.5% Triton X-100, and 2 mg/ml lysozyme. Purification is achieved by ammonium sulfate fractionation followed by sizing, ion exchange and hydroxyapatite chromatography. Details of the purification and functional assays of the resulting fusion protein are described elsewhere (D.L.R., J.H. Sinard and T.D.P., in preparation). We have used the purified fusion proteins to study the role of the myosin tail in polymerization. The resulting fusion protein is functional in that its assembly properties are equivalent to those of native myosin-II.

(c) Conclusions

Though expression of eukaryotic proteins in *E. coli* is used frequently, the process is not yet well characterized. It is clear that the conformation and amount of the resulting protein are a function of the sequence used. Given this variable, the investigator must try different vectors and host systems to obtain the most favorable results (see Leinwand et al., 1988 or Denhardt and Colasanti, 1987 for review). We have described a new vector system that, when used with a protease deficient host, yields large amounts (100 mg/liter or 45% of the soluble protein) of functional myosin tail fusion protein. We hope that this system will be valuable for the bacterial expression of other eukaryotic proteins when large amounts of fusion protein are desired.

ACKNOWLEDGEMENTS

We thank Dr. Don Cleveland for providing the pATH vectors and CAG-456 strain and Dr. L. Leinwand for a preprint of her review of expression of contractile proteins in *E. coli*. Thanks also to Drs. T.J. Koerner and Don Cleveland for their suggestions and critical reading of the manuscript. This work was supported by NIH research grant GM-26132 and a predoctoral fellowship to D.L.R. from the Johns Hopkins School of Medicine Medical Scientist Training Program (GM07309).

REFERENCES

Bachmair, A., Finley, D. and Varshavsky, A.: In vivo half-life of a protein is a function of its amino terminal residue. Science 234 (1986) 179–186.

Baker, T.A., Grossman, A.D. and Gross, C.A.: A gene regulating the heat shock response in *Escherichia coli* also affects proteolysis. Proc. Natl. Acad. Sci. USA 81 (1984) 6779–6783.

Courtney, M., Buchwalder, A., Tessier, L.-H., Jaye, M., Benavente, A., Balland, A., Kohli, V., Lathe, R., Tolstoshev, P. and Lecocq, J.-P.: High level production of biologically active human α-antitrypsin in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81 (1984) 669–673.

Denhardt, D.T. and Colasanti, J.: A survey of vectors for regulating expression of cloned DNA in *E. coli*. In Rodriguez, R.L. and Denhardt, D.T. (Eds.) Vectors. Butterworths, Stoneham, MA, 1987, pp. 179–204.

Dieckmann, C.L. and Tzagoloff A.: Assembly of the mitochondrial membrane system. J. Biol. Chem. 260 (1985) 1513–1520.

Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F. and Cleveland, D.W.: Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. J. Cell Biol. 104 (1987) 817–829.

Goff, S.A. and Goldberg, A.L.: Production of abnormal proteins in *E. coli* stimulates transcription on *lon* and other heat shock genes. Cell 41 (1985) 587–595.

Hare, D.L., Sadler, J.R. and Betz, J.L.: Regulated high-level expression of the *Herpes simplex* type I thymidine kinase gene in *Escherichia coli*. Gene 32 (1984) 117–128.

Harms, T.J.R.: Expression of eukaryotic genes in *E. coli*. In Williamson, R. (Ed.), Genetic Engineering, 4th Ed. Academic Press, New York, 1983, pp. 128–185.

Laemmli, U.K.: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227 (1970) 680–685.

Leinwand, L.A., Sohn, R., Frankel, S.A., Goodwin, E.B. and McNally, E.M.: Bacterial expression of eukaryotic contractile proteins. Cell Motil. Cytoskelet. (1988) (in press).

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

Putkey, J.A., Slaughter, G.R. and Means, A.R.: Bacterial expression and characterization of proteins derived from the chicken calmodulin cDNA and a calmodulin processed gene. J. Biol. Chem. 260 (1985) 4704–4712.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Straus, D. and Gilbert, W.: Chicken triosephosphate isomerase complements an *Escherichia coli* deficiency. Proc. Natl. Acad. Sci. USA 82 (1985) 2014–2018.

Vieira, J. and Messing, J.: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19 (1982) 259–268.