J. Biosci., Vol. 21, Number 5, September 1996, pp 687–697. © Printed in India.

Mutagenesis study to ameliorate the bacterial expression of phosphoprotein P of vesicular stomatitis virus

SUSANTA PAHARI and D J CHATTOPADHYAY*
Department of Biochemistry, University of Calcutta, 35, Ballygunge Circular road, Calcutta 700 019, India

MS received 16 February 1996; revised 4 June 1996

Abstract. The phosphoprotein gene of vesicular stomatitis virus, a Rhabdovirus, has been inserted into bacterial expression plasmids containing the *Escherichia coli tac* promoter and ribosome binding site (RBS). A low level of expression of the protein was detected. Sequence analysis showed the presence of 15 nucleotides in the spacer region i.e., between the Shine-Dalgarno sequence and ATG. Alteration of the distance and the sequence in the spacer region by oligonucleotide-directed mutagenesis revealed a correlation among the expression levels, accessibility of the RBS and requirement for a minimum spacing of at least 7 nucleotides between the Shine-Dalgarno sequence and ATG for optimal gene expression.

Keywords. VSV P protein; overexpression; parameter optimization; 5′untranslated region; mutagenesis.

1. Introduction

Vesicular stomatitis virus (VSV), a prototype of Rhabdoviridae family, needs the help of mainly two proteins, L and P, for its transcription *in vivo* as well as *in vitro* (Banerjee 1987). While the L protein carries out the transcription of the genome and subsequent modifications of the transcripts the P protein acts as a transcriptional activator. *In vitro* transcription experiment has established that L is needed in catalytic amount whereas P is required in stoichiometric amount (De and Banerjee 1985). The P protein has been found to be a phosphoprotein (Clinton *et al* 1979) and the extent of phosphorylation has a profound effect on transcriptional activity (Kingsford and Emerson 1980). The more highly phosphorylated form(s) generally shows greater activity and treatment with phosphatase causes inactivation (Masters and Banerjee 1986). Mutation of two specific serine residues in P protein domain II, at position 236 and 242, reduces the activities of the P protein indicating that phosphorylation of these two serine residues is very much essential for the transcriptional activation (Chattopadhyay and Banerjee 1987). The aminoterminal domain 1 of P protein was also found to be phosphorylated by host kinase (Barik and Banerjee 1991). To find out the protein kinase(s) responsible for the phosphorylation of P and subsequently to ware out the exact mechanism underlying the transactivation of the transcription reaction it is necessary to overexpress the P proteins, in a suitable system. We decided to overexpress the P protein in a prokaryotic system *viz.*, *Escherichia coli* because prokaryotic kinase(s) are highly specific and less abundant in comparison to eukaryotes. Therefore it may be possible to obtain the expressed P in a dephosphorylated form. The posibility of purifying P from

*Corresponding author (Fax, 091-033-241-3222; Email: dhruba@cubmb.ernet.in).

the virus source-has been excluded considering the original viral source too tiny to yield a significant amount of pure P. Additionally, the viral source will give phosphorylated P and dephosphorylation with the help of bacterial alkaline phosphatase may not remove all phosphates which remains internalized due to the formation of secondary and tertiary structure. Earlier, Barik and Banerjee (1991) have expressed the phosphoprotein in P in *E. coli* using T7 promoter system at an extremely high level. But the high level expression results in the accumulation of P protein in inclusion bodies. Therefore to avoid the formation of inclusion body, expression using a relatively weak promoter, *tac*, was envisaged.

Two *tac* promoter vectors viz., pKK223-3 (Pharmacia) and pJF118EH (Furste *et al* 1986) were employed in the present work. *tac* promoter is a hybrid promoter containing – 10 region of the lac UV5 promoter and – 35 region of the trp promoter. It is five times stronger then the lac UV5 promoter (Deboer *et al* 1982). The difference between these two vectors is the absence of a 'lac' repressor overproducing sequence ($laci^q$) in the former but it's presence in the latter. Results showed that the level of expression is low but among the two. pJF118EH is better for overexpression purpose. To find out the reasons behind the low level expression in pJF118 P(NJ), emphasis had been laid upon translation parameters. Stormo (1986) had developed a consensus rule to achieve optimum level of translation. This rule suggests that efficient initiation will require an optimal SD sequence positioned about seven to nine nucleotides upstream of the initiator codon. This information prompted us to reduce the number of nucleotides between SD sequence and ATG. Four mutants have been constructed by (i) employing oligo-directed deletion mutagenesis and (ii) enzymatically removing an unique *Eco*RI site in the spacer region. Then computer analysis for the secondary structure from – 20 to + 20 bp region relative to the AUG translation initiation codon was done and lastly, induction, SDS-PAGE analysis followed by Western blotting was performed to observe the change in the level of expression.

2. Materials and methods

2.1 *Cloning of P(NJ) simultaneously into pKK223-3 and pJF118EH*

The P(NJ) fragment was available from pGEM4 P(NJ) (Gill *et al* 1986). pGEM4 P(NJ) plasmid DNA was, digested with *Eco*RI and *Hind*III. The required P(NJ) fragment was eluted from 0·8% agarose gel and was used to ligate with *Eco*RI- and *Hind*III-digested pKK223-3 and pJF118EH. The ligate DNAs were used to transform the competent *E. coli* cells. The positive recombinants, pKK223-3 P(NJ)XLIBlue and pJF118 P(NJ) XLIBlue were selected by *Eco*RI and *Hind*III digestion of the isolated plasmid DNAs.

2.2 *Induction, SDS-PAGE analysis and Western blotting*

Recombinant positive clones were induced by isopropyl-*β*-D-thio-galactopyranoside (IPTG), an inducer for *tac* promoter expression. Cells were harvested, lysed by SDS-PAGE sample buffer and loaded onto a 10% SDS-PAGE in duplicate. One part was subjected to Coomassie blue staining and another part transferred to nitro-cellulose membrane and probed with P(NJ) antisera. To measure the signal, chemiluminescent assay procedure (BIORAD) was performed.

2.3 *Standardization of physiological parameters to optimize the expression*

Five physiological parameters viz·, host stain, medium, optical density of the culture during addition of inducer, inducer concentration and induction range were standardised. Nine different strains of different genotype like BH101, D1210 (HB101, laci^q), SG117 (HB101, lon$^-$), SURE (laci^q), JM109 (laci^q), XLIBlue (laci^q), BL21 (OmpT$^-$), CAG456 (lon$^-$ htpr 165) had been tested. Two media viz., Luria broth (LB: bacto-tryptone, 10 g/litre; bacto-yeast extract, 5 g/litre; NaCl, 10 g/litre) and Terrific broth (TB: bacto-tryptone, 12 g/litre; bacto-yeast extract, 24 g/litre; glycerol, 4 ml/litre, 0·17 M KH_2 PO_4, 0·72 M K_2 HPO_4 (Sambrook *et al* 1989) were used. The cells were induced at an optical density at 600 nm (OD_{600}) of 0·3, 0·5 and 0·7. IPTG concentration of 0·1 mM, 0·2 mM, 0·5 mM and 1 mM were used in different sets. Induction period of 2 h, 4 h, 6 h and 8 h was tested.

2.4 *Mutagenesis*

Deletion between Shine-Dalagarno sequence (AGGA) and translation initiation codon (ATG) was done by oligo-directed mutagenesis and by restriction enzyme digestion.

2.4a *Oligo-directed deletion mutagenesis*: The mutagenesis was carried out by following the method of Kunkel *et al* (1987). Deletions were performed by removing 4 nucleotides at a time giving rise to mutated spacer length of 11, 7 and 3 nucleotides.

2.4b *Design of oligos*: Wild type sequence: 5′ TTCACAC (AGGA) AACAGAA-TTCCCATC (ATG) GACA 3′, Mutant type (A) PRP16 (30 mer): 5′ TGTC (CAT) GATGGGAAGTT (TCCT) GTGTGAAA 3′; (B) PRP17 (30 mer): 5′ ACACTGTC (CAT) GATGGTT (TCCT) GTGTGAA 3′; (C) PRP18 (30 mer): 5′ATCAACACTGTC (CAT) GTT (TCCT) GTGTGAAA 3′. All these oligos were synthesized in an oligo synthesizer (Model 380B DNA synthesizer, Applied Biosystem), deprotected and 5′ ends were phosphorylated with T4 polynucleotide kinase.

2·4c *Subcloning of* [*ptac and P*(*NJ*)] *fragment into pBSKS*: P(NJ) alongwith *tac* promoter sequence (*ptac*), available from pKK223-3 P(NJ) by *Bam*HI digestion was subcloned into bluescript phagemid pBSKS(+). pBSKS(+) was chosen to prepare single stranded uracil containing DNA for mutagenesis by Kunkel's method. The resultant recombinant plasmid had been designated as pBS*Bam*HIP(NJ). This clone contains the spacer region (AGGA to ATG) where mutagenesis has to be done.

2.4d *Single stranded uracil containing DNA preparation*: pBS*Bam*HI P(NJ) DNA was transformed into RZ1023. The pBS*Bam*HIP(NJ)RZ1023 cells were allowed to grow up to mid log phase in 2 × Ty broth containing amp (100 μg/ml) and tet (10 μg/ml) using 1% inoculum from overnight grown culture. At mid log phase, R4108 helper phage (0·02–0·2 pfu/cell) and uridine (0·25 μg/ml) were added and allowed to grow for 6 h at 37°C with vigorous shaking. Supernatant which contains phages were recovered by centrifugation at 5000 *g* for 30 min at 4°C (Hitachi, CR20B2). To the phage suspension, 0·25 vil of NaCl/PEG solution (15% w/v PEG 8000 and 2·5 M NaCl) was added, kept on ice for 1 h and centrifuged at 5000*g* for 20 mM at 4°C. The phage pellet was

resuspended in TE (*p*H 7·6) and again centrifuged at 5000 *g* for 20 min at 4°C. Then the pellet was extracted twice with phenol (saturated with Tris-HCI pH 8·0) and once with phenol: choloroform: isoamyl alcohol (25:24:1). Aqueous phase from the final extraction was transferred to a glass centrifuge tube, ethanol precipitated in the presence of 0·1 vol of 3 M Na-acetate (pH 4·8), washed with 70% ethanol, dried and dissolved in TE (pH 7·7) (Sambrook *et al* 1989).

2.4e *Mutagenesis reaction*: ss-uracil containing DNA and 5′ end phosphorylated oligos were annealed in 20 mM Tris-HCI pH 7·5, 10 mM $MgCl_2$, 50 mM NaCl, 1mM DTT, for 5 min at 60°C, slowly cooled down to RT, extended with sequenase (Schena 1989) and ligated using T4 DNA ligase.

2.4f *Screening of mutants*: Mutants were selected by *Bam*HI digestion and *Eco*RI digestion. *Bam*HI digestion yielded a fragment of 1132 bp whereas *Eco*RI digestion resulted only in linearization because *Eco*RI site of the mutant was lost. Finally sequencing was done for conformation. Mutant clones corresponding to oligo PRP16, PRP17, PRP18 are designated as pBSPRP16, pBSPRP17, pBSPRP18 respectively.

2.4g *Subcloning of mutants into pRCP-5*: pRCP-5 is a new construct derived from pJF118. For examining the effect of deletion upon expression in absence of wild type *tac* promoter, it was removed from pJF118. The detailed procedure is shown schematically in figure 1. Restriction digestion and modifications required for the *tac* promoter removal created simultaneously a new unique *Bam*HI site in pRCP-5. This newly created unique *Bam*HI site facilitated subsequent subcloning of mutated fragments. pBSPRP16, pBSPRP17, pBSPRP18 ad pKK223-3 were digested with *Bam*HI, inserts were purified and used for ligation with *Bam*HI digested pRCP-5, transformed to XLIBlue, selected by *Bam*HI digestion and orientation checked by NarI digestion. The resultant mutant clones corresponding to pBSPRP16, pBSPRP17, pBSPRP18 and pKK223–3 P(NJ) in pRCP-5 are designated as M1, M2, and M3 and wild type (W).

2.4h *Enzymatic deletion mutagenesis of pJF118P*(*NJ*): Sequence shown between 8 bp upstream of SD sequence and 4 bp downstream of translation initiation codon (ATG) was 5′ TTTCACAC(AGGA)AAC (GAATTC)CCATC(ATG)CACA 3′. Therefore, digestion with *Eco*RI and removal of overhang nucleotides will delete 4 nucleotides (AATT), giving rise to a new mutant of 11 nucleotides spacer length but having different composition with respect to M1. Accordingly, pJF118P(NJ) digested with *Eco*RI, treated with mung bean nuclease (MBN) ad subjected to ligation by T4DNA ligase. Ligation mix was transformed to XLIBlue and selected by *Eco*RI digestion. Clones that have lost the *Eco*RI site were selected. Subsequently sequencing was done to confirm the resultant mutated sequence: 5′TTTCACAC (AGGA)AAC-GCCCATC(ATG)CACA3′, designated as M4.

2·4i *Secondary structure analysis*: A computer aided analysis of the nucleotide sequence of the wild and mutant types allowed us to construct a theoretical secondary structure around the initiator AUG and its free energy values. This was done using the program 'FOLD' (Devereux *et al* 1984) available in GCG package at Distributed Information Centre (DIC), Bose Institute, Calcutta.

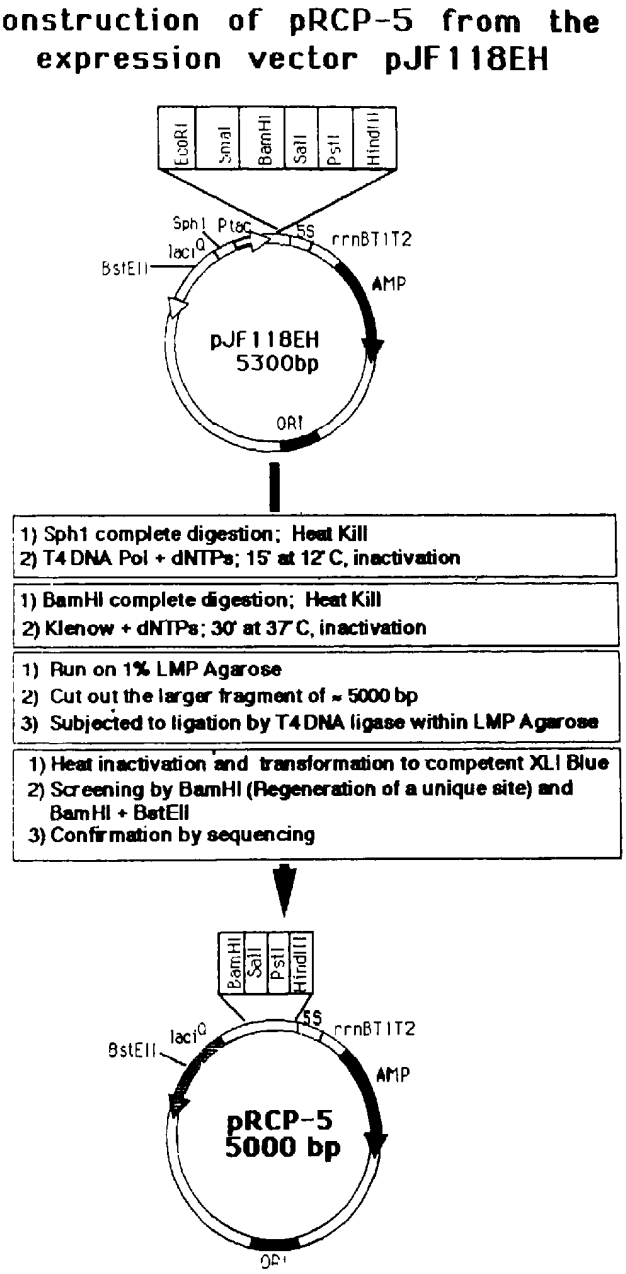

Figure 1. Construction of pRCP-5 from the expression vector pJF118EH.

3. Results

In an attempt to overexpress Pin *E. coli* we had started with a very simple system using pKK223–3 and pJF118EH simultaneously.

3.1 *Standardization of physiological parameters for optimal expression*

Optimization of physiological parameters responsible for expression was examined. The yield of P protein was found to be similar in all the *E. coli* strains used. So XLIBlue was used for all subsequent experiments. Terrific broth (TB) being an enriched medium

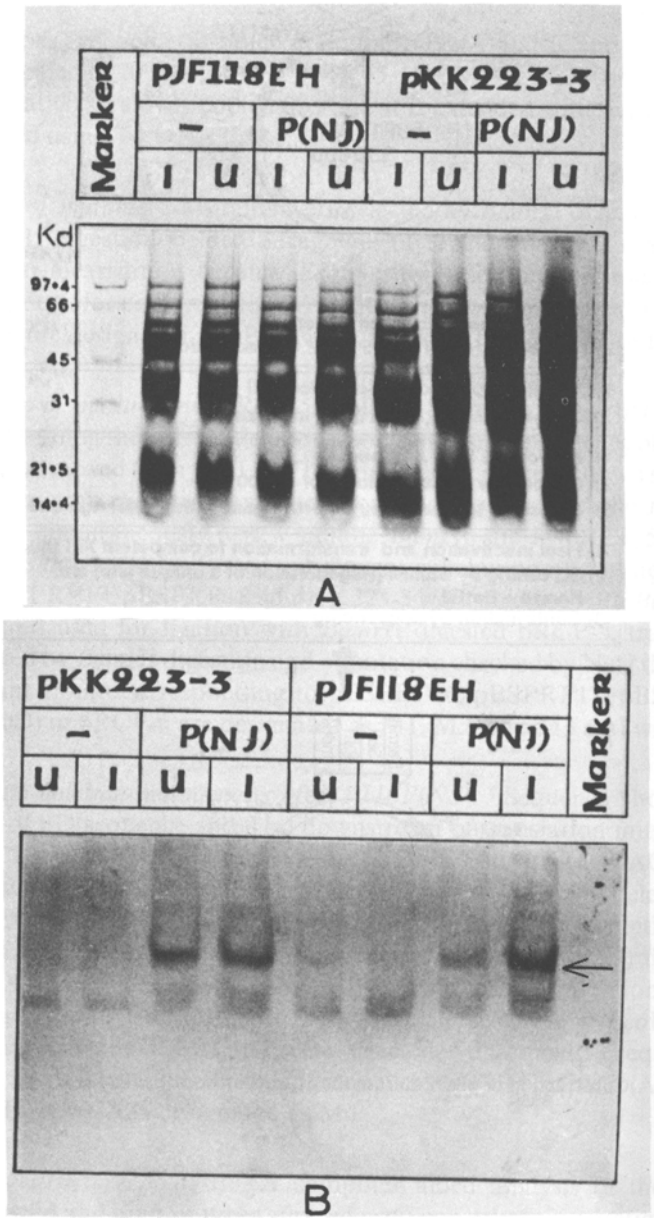

Figure 2. (**A**) Expression profile of pKK223–3P(NJ)XLIBlue and pJF118P(NJ)XLIBlue. I and U denote induced and uninduced lanes. (**B**) Western blotting picture of the gel shown in (**A**). Lanes are denoted as in the stained gel

was found to be better. OD_{600} of 0·5 was found to be the best with respect to protein induction and 0·5 mM IPTG concentration gave best result. Up to 8 h of induction period was found to be optimal.

3.2 *Expression profile of pKK223–3P(NJ) XLIBlue and pJF118P(NJ) XLIBlue*

After fixing all these parameters, the cultures were induced and protein was analysed. A 10% SDS-PAGE profile and Western blotting of this experiment had been shown in figure 2. Although no new band was seen in induced cells Western blotting with antibody gave a distinct signal at 49 kDa, confirming the expression of P protein. The expression in the uninduced lane of pJF118 P(NJ) XLIBlue was almost negligible whereas in case of pKK223–3P(NJ) XLIBlue, both the uninduced and induced lanes showed almost equal amounts of P protein (figure 2). Thus the presence of $laci^q$ sequence in the pJF118 P(NJ) conferred a good repression before induction. So pJF118 which had been found to be superior was used for subsequent experiments.

3.3 *Salient features of the secondary structure of mutants and wild type*

The computer analysis for the secondary structure of the four mutants was done alongwith the original construct. Results obtained are shown in table 1. The analysis was done from – 20 to + 20 bp region relative to the AUG. Salient features appeared for the mutants were: (i) M1 and M4 each has a spacing of 11 nucleotide but different nucleotide composition. (ii) M4 had the most stable structure with free energy of – 7·9 Kcal whereas M3 had the least stable structure with free energy of – 1·1 Kcal. (iii) Mutants M1, M2 and also wild type, W, possessed almost similar free energy of – 3·1, – 3·2 and – 3·4 Kcal respectively. (iv) M2 had a looped SD sequence but the other mutants and wild type SD sequence was free.

3.4 *Expression level determination of wild type and mutants*

Wild type and four mutants (M1, M2, M3, M4) were induced, pellet lysed, equal amounts of protein were loaded on a 7·5% SDS-PAGE in duplicate. One part was used for Coomassie staining and another part subjected to Western blotting (figure 3). The level of P protein synthesized was measured by scanning the radio autographs. Densitometric scanning analysis showed that the level of expression in M1 > W > M2 > M4 > M3.

4. Discussion

In an initial attempt to overexpress P in *E. coli* we had started with a very simple system using pKK233–3 and pJF118EH simultaneously. It was that expression from pKK233–3P(NJ) occurs almost at equal level both in uninduced and induced conditions whereas for pJF118P(NJ), the basal level expression in induced condition was low in comparison to induced condition. This proved that use of both $laci^q$ strain and plasmid borne $laci^q$ gene as source of lac repressor enabled a good basal repression.

Parameters influencing the efficiency of expression of the P(NJ) gene in *E. coli* were studied by comparing the expression of P from four *in vitro* derived spacer length variants (M1, M2, M3, M4). Computer model showed that in the two mutants, M3 and M4 SD sequence remains free indicating its availability to ribosome and consequently removal of high energy barrier has been possible only to a limited extent. M2 has a looped SD sequence which inhibited accessibility of ribosome to mRNA thus giving rise to low yield. But 7 nucleotides spacer length and low energy values, in combination, made moderately favourable circumstances for expression. Probably, the looped SD sequences is mutable due to its low energy values and once melted, 7 nucleotides spacer length favoured translational initiation. Thus looped SD sequence barrier had been overcome. Original sequence showed better expression than former three mutants (M2, M4 and M3). This happened due to its free SD sequence and having comparable stability (– 3·1 Kcal) with others (table 1). M1 showed the highest level of expression

Table 1. Salient features of the wild and mutant types spacer region.

Sequence	No. of spacing nucleotides	Secondary structure	Free energy (Kcal)	Status of SD sequence
W	15	10 20 AGGAAACA AUU C UGGA GA CC AUCA C CU GG UAGU A -----GAA --C A UGUG 40 30	– 3·4	Free
M1	11	10 20 AGGAAA --C – UGGA CUU CC AUCA C GAA GG UAGU A ------ CUC A UGUG 30	– 3·1	Free
M2	7	10 –AG AA – UGGA GA CC AUCA C CU GG UAGU A GAA –C A UGUG 30 20	– 3·2	Looped
M3	3	10 -----AGGA GG AACAU UUGUG A GAACUCGGAUAG AC 30 20	– 1·1	Free
M4	11	10 20 AGGAAAC C UGGA AGCC AUCA C UCGG UAGU A ---GAAC A UGUG 30	– 7·9	Free

The experimental details are as in §2.

because this has all the criteria owned to wild type, additionally it has a spacer length of 11 nucleotides. This indicated that 11 nucleotides appeared more favourable than a long distance of 15 nucleotides of wild type for P expression. M1 ad M4 have 11 nucleotides spacer length each but have a different composition. Most interesting is the

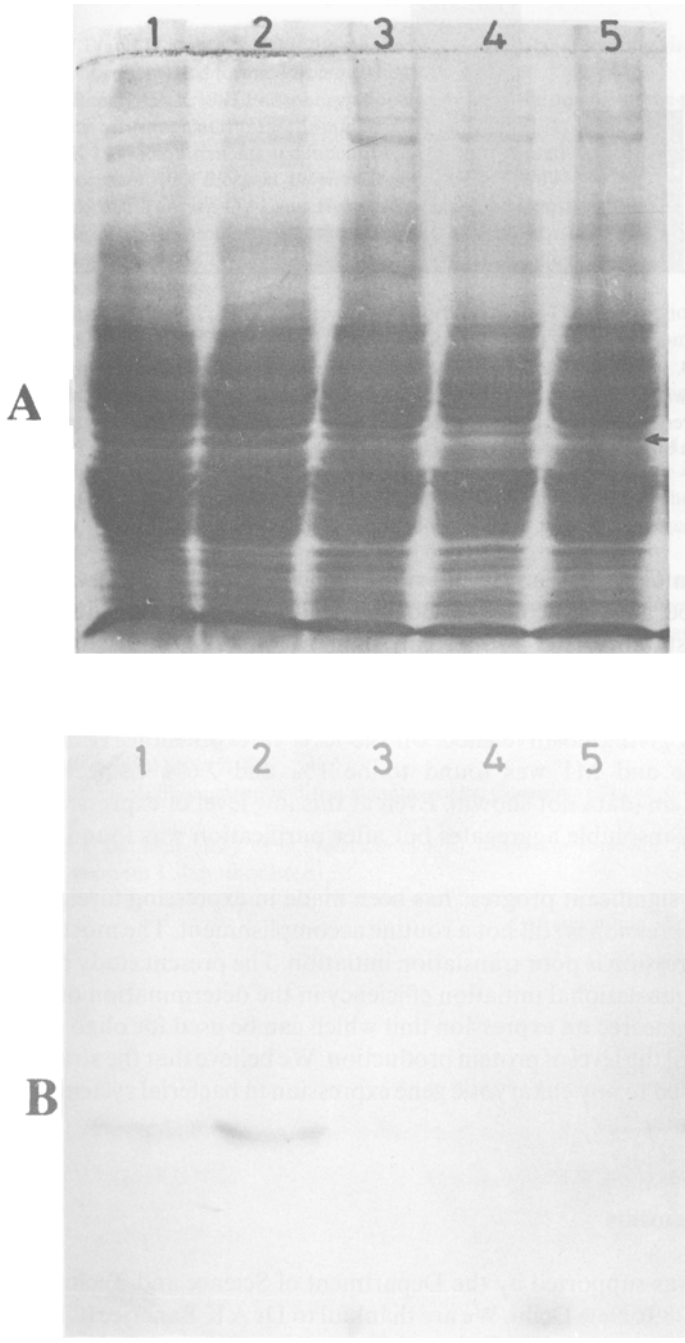

Figure 3. (**A**) and (**B**)

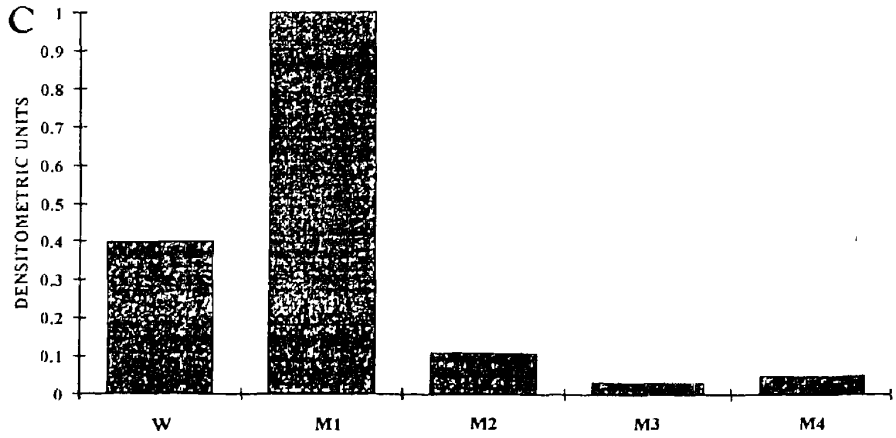

Figure 3. 7·5% SDS-PAGE profile of wild type and mutants after inducing with IPTG. (**A**) Lane 1, wild type; lane 2, M1; lane 3, M2; lane 4, M3; lane 5, M4. (**B**) Western blotting picture of the corresponding gel. Lanes are denoted as in stained gel. (**C**) Densitometric scanning of Western blot.

presence of a G residue in M4 but not in M1. In M1, the region of G poor sequence proved to be advantageous (de Boer *et al* 1983). The G residue in M4 provided a stronger stability (ΔG – 7·9 Kcal) than M1 (ΔG – 3·1 Kcal). Therefore higher level of expression in M1 was obtained. The M2 and M3 have a long stem structure, whereas M1 and wild type have short stem structures which is easy to be melted, thus giving positive effect on the level of expression. The level of P protein in wild type and M1 was found to be 1% and 2·6% respectively of the total cellular protein (data not shown). Even at this low level of expression the protein was recovered as insoluble aggregates but after purification was found to be biologically active.

Although significant progress has been made in expressing foreign genes in *E. coli*, high level expression is still not a routine accomplishment. The most common block to efficient expression is poor translation initiation. The present study has highlighted the key role of translational initiation efficiency in the determination of expression levels. We have engineered an expression unit which can be used for oligo directed mutagenesis to control the level of protein production. We believe that the strategy outlined here can be adapted to any eukaryotic gene expression in bacterial system under the control of *tac* promoter.

Acknowledgements

This work was supported by the Department of Science and Technology (Grant No. SP/SO/D75/89), New Delhi. We are thankful to Dr A K Banerjee (Cleveland, Ohio) for pGEM4P(NJ) clone and Dr Robert E Shope (New Haven, CT) for P (NJ) antisera and DIC, Bose Institute, Calcutta, for computer facility.

References

Banerjee A K 1987 Transcription and Replication of Rhabdoviruses; *Microbiol. Rev.* **51** 66–87

Bulk S and Banerjee A K 1991 Cloning and expression of the vesicular stomatitis virus phosphoprotein gene in *Escherichia coli*: analysis of phosphorylation status versus transcriptional activity; *J. Virol.* **65** 1719–1726

Clinton G M, Burge W B and Huang A S 1979 Phosphoprotein of vesicular stomatitis virus: identity and interconversion of phosphorylated forms; *Virology* **99** 84–94

Chattopadhyay D and Banerjee A K 1987 Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription *in vitro*; *Cell* **49** 407–414

De B P and Banerjee A K 1985 Requirements and functions of vesicular stomatitis virus L and NS proteins in the transcription process *in vitro*; *Biochem. Biophys. Res. Commun.* **126** 40–49

de Boer H A, Comstock L J, Yansura D G and Heynecker H L 1982 Construction of a tandem trp-lac promoter and a hybrid trp-lac promoter for efficient and controlled expression of the human growth hormone gene in *Escherichia coli*; in *Promoters, structure and function* (eds) R L Rodriguez and M J Chamberlin (New York: Praeger) pp 462–481

de Boer H A, Hui A, Comstock L J, Wong E, Vasser M 1983 Portable Shine-Dalgarno regions: a system for a systematic study of defined alterations of nucleotide sequences within *E. coli* ribosome binding sites; *DNA* **2** 231–235

Devereux J, Haeberli P and Smithies O 1984 A comprehensive set of sequence analysis programmes for the VAX; *Nucleic Acids*. *Res*. **12** 387–395

Fürste J, Pansegrau W, Frank R, Helmut B, Scholz P, Bagdasarian M and Lanka E 1986 Molecular cloning of the plasmid RP4 primase region in a multi-host range tacP expression vector; *Gene* **48** 119–131

Gill D S, Chattopadhyay D and Banerjee A K 1986 Identification of a domain within the phosphoprotein of vesicular stomatitis virus that is essential for transcription *in vitro*; *Proc. Natl. Acad. Sci. USA* **83** 8873–8877

Kingsford L and Emerson S U 1980 Transcriptional activities of different phosphorylated species of NS protein purified from vesicular stomatitis virions and cytoplasm of infected cells; *J. Virol.* **33** 1097–1105

Kunkel T A, Roberts J D and Zakour R A 1987 Rapid and efficient site-specific mutagenesis without phenotypic selection; *Methods Enzymol*. **154** 367

Masters P S and Banerjee A K 1986 Phosphoprotein of vesicular stomatitis virus: phosphorylated states and transcriptional activities of intracellular and virion forms; *Virology* **154** 259–270

Stormo G 1986 Translation init1ation; in *Maximizing gene expression* (eds) W Reznikoff and L Gold (Boston, Massachusetts: Butterworth) pp 195–224

Schena M 1989 High efficiency oligonucleotide-directed mutagenesis; *Comments* (*United States Biochemical Corp.*) **15** 23

Sambrook J, Fritsch E F and Maniatis T 1989 *Molecular cloning*, *a laboratory manual* 2nd edition (New York: Cold Spring Harbour Laboratory Press)

Corresponding editor: M S SHAILA