# An insight into the possible mechanism of working of two-cistronic gene expression systems and rational designing of newer systems

UTPAL KUMAR MUKHOPADHYAY* and GIRISH SAHNI†

*Institute of Microbial Technology, Sector 39-A, Chandigarh 160 036, India*

**Present address: Molecular Biology Unit, Animal Biotechnology Centre, National Dairy Research Institute, Karnal 132 001, India*

†*Corresponding author (Fax, 91-172-690585; Email, sahni@imtech.ernet.in)*

The initial attempts at hyper-expressing buffalo/goat growth hormone (GH)-ORFs in *Escherichia coli* directly under various strong promoters were not successful despite the presence of a functional gene. High level expression of GH was achieved as a fusion protein with glutathione-S-transferase (GST). To produce native GH in an unfused state, we adapted an established strategy of two-cistronic approach in our system. In this strategy, utilizing one of the highly efficient reported sequences as the first cistron led to a nearly 1000-fold enhancement in the level of expression under an *E. coli* promoter (*trc*). In search of a newer first-cistron sequence as well as to see the generality of the two-cistronic approach, we explored the ability of different lengths of a highly expressing natural gene to act as an efficient first cistron. Surprisingly, *GST*, which is naturally highly expressible in *E. coli*, could not be fitted into a successful two-cistronic construct. In addition, placement of the entire two-cistronic expression cassette (which had earlier given high-level GH expression under *trc* promoter) under the *T7* promoter in *E. coli* failed to hyper-express GH. These results suggest that the successful exploitation of the two-cistron arrangement for hyper-expression of eukaryotic ORFs in bacteria is not as straightforward as was previously thought. It appears probable that factors such as the sequence context, together with the length and codons used in the first cistron are important as well.

[Mukhopadhyay U K and Sahni G 2002 An insight into the possible mechanism of working of two-cistronic gene expression systems and rational designing of newer systems; *J. Biosci*. **27** 219–231]

## 1. Introduction

Translational initiation events have been seen to be a limiting factor for achieving hyper-expression in *E. coli* of many genes of eukaryotic origin, particularly those bearing extensive complementarity between the 5′-untranslated region and the 5′-coding sequence which forms strong and stable secondary structure/s through intramolecular base pairing. Such structures, especially those involving the ribosome binding site (RBS = SD + ATG) are believed to render them relatively inaccessible to the ribosome, leading to translational repression (e.g. de Smit and van Duin 1994; Wang *et al* 1995; Linder and Tuite 1999; van Marteen *et al* 2001). To overcome the potential problem of translational repression due to mRNA secondary structures, an elegant strategy based on a synthetic two-cistronic expression system that drastically de-represses the translation has been described (Schoner *et al* 1986). In this system, a short synthetic cistron (cistron 1) is placed upstream of the gene sought

**Keywords.** *Escherichia coli*; gene expression; growth hormone; mRNA secondary structure; translational de-repression; two-cistron

Abbreviations used: GH, Growth hormone; GST, glutathione-S-transferase; INCE, internal negative control element; RBS, ribosome binding site; RRF, ribosome releasing factor.

to be expressed in such a way that the stop codon for the upstream cistron (stop 1) falls within the RBS (between SD and ATG) of the downstream cistron (cistron 2). It has been proposed that when the ribosomes translating the first cistron reach the stop codon, they probably facilitate the translation of the downstream cistron by 'melting' the hindering secondary structure/s (Schoner *et al* 1987). However, the underlying mechanism by which the expression of the downstream cistron is improved through such an arrangement has not been rigorously tested and remains to be experimentally checked. Similar studies carried out with natural *Escherichia coli* polycistronic transcripts (Das and Yanofsky 1984; Wikstrom *et al* 1992) showed that merely relocating the stop codon close to the downstream initiation site drastically improves translation of an otherwise poorly expressed cistron. From these studies, the mechanism by which a terminating ribosome entering at upstream RBS (i.e. RBS1) would improve the overall translational efficiency of the downstream cistron was thought to be due to one or more of the following factors:

(i) The same 70S ribosome can 'read through' the cistron 2 after reading cistron 1.
(ii) The terminating ribosome physically disrupts secondary structures, thereby exposing SD2 and ATG2, thus allowing efficient entry of fresh ribosome at RBS2.
(iii) The increased local concentration of ribosomal subunits around RBS2 due to terminated ribosomes leads to an increased rate of translation initiation at RBS2.

The present study indicates, however, that additional factors related to 'compatibility' between the two cistrons can also be important in the successful design of the two-cistron expression systems. These studies were prompted by our observation that the *GST* gene, which is naturally highly expressible in *E. coli*, could not be successfully utilized to derive a two-cistronic construct despite meeting with all other previously suggested design requirements of the two-cistron expression systems. The results of our study are reported below.

## 2. Materials and methods

Most commonly-used methods e.g. restriction endonuclease digestion, kinasing, T4 DNA polymerase-mediated end-polishing, DNA ligation, bacterial transformation etc. were carried out as described (Sambrook and Russell 2001). Other methods specific to this study are described briefly with appropriate modifications wherever applicable. The nucleotide sequencing of all the newly developed constructs were carried out by Sanger's method (Sanger *et al* 1977) to confirm faithful incorporation of desired alterations.

### 2.1 *Bacterial strains and plasmids*

Plasmid vectors pTrc99A (Amann 1988), pET23d$^+$ (Studier *et al* 1990), pGEX-KG (Guan and Dixon 1991) and pGEX-KT (Hakes and Dixon 1992) were used to construct different GH-expression vectors. For initial cloning and characterization of constructs as well as for expression studies using *E. coli* derived promoters, *E. coli* strain XL1 Blue (Stratagene Inc., CA, USA) was used. The *E. coli* strain BL21 (*DE3*) (Novagen Inc., WI, USA) was used for studying T7 promoter-based expression.

### 2.2 *Construction of a synthetic two-cistron expression system for* GH *gene/s under* trc *and* T7 *promoters*

Two synthetic oligonucleotides BGHB1 (5′-CATGGAG GGTATTAATAATGTATCGATTAAATAAGGAGGAA TAACA-3′) and BGHB2 (5′-TATGTTATTCCTCCT TATTTAATCGATACATTATTAATACCCTC-3′) were designed, so that upon annealing, they would form a short synthetic cistron with a 8-amino acid coding sequence ending with a stop codon. This synthetic cistron also contained a RBS (i.e. RBS2) located at the 3′-end of the cistron (figure 1). This downstream RBS is designed to be used by the second cistron (*GH* cDNA). The *GH*-ORF was spliced downstream to the synthetic cistron and was subcloned under the *trc* promoter in the vector pTrc99A to develop pUG99IIB.

The entire two-cistronic cassette was taken out on a *Nco*I-*Bam*HI fragment and subcloned between *Nco*I and *Bam*HI sites in pET23d$^+$ (to develop pUGET23IIB), the *T7* promoter based vector. This construct was initially transformed into *E. coli* XL1-Blue for characterization, following which it was retransformed into *E. coli* strain BL-21 (*DE3*) for GH expression.

### 2.3 *Construction of GST-GH two-cistronic expression system*

Two-cistronic expression cassettes were also constructed using different lengths of *GST* gene as the first cistron and the *GH* gene as the second cistron in the pGEX-KT/pGEX-KG vector (figure 2). In the expression vector pMSEX-2G (figure 2b) the full length *GST* gene was used as the first cistron, where a translational stop codon at the end of the *GST* gene as well as a RBS in the beginning of the *GH* gene was introduced through a pair of synthetic oligonucleotides UKM-1 (5′-GGGGGAGGAA TAACC-3′) and UKM-2 (5′-GGTTATTCCTCCCCC-3′).

Using a truncated version of *GST* as first cistron, two more two-cistronic expression vectors namely pMSEX-2G8 and pMSEX-2G16 (figure 2c, d), were also constructed using synthetic oligonucleotides UKM-3 (5′-

ACTAGGGGAGGAATAACC-3′) and UKM-4 (5′-C ATGGGTTATTCCTCCCCTAG-3′) and UKM-5 (5′-ACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAG GAGGAATAACC-3′) and UKM-6 (5′-CATGGGTTAT TCCTCCTGCACAAGGCCCTTAATTTTCCAATAACC TAG-3′), respectively. In both cases, the first cistron encoded a short peptide of either 8 or 16 amino acids (including initiator codon), respectively. In all the *GST*-based two-cistronic constructs, the SD sequence, the distance between SD and ATG of the downstream RBS as well as positioning of the translational stop codon for the first cistron were kept exactly similar to that of construct pUG99IIB.

2.4a *Detection of gene expression: growth of cells and induction of gene expression*: Inoculum was raised by seeding 5 ml LB-Ampicillin (100 µg/ml) with a single, well-isolated colony (approximately 2 mm in diameter) from a freshly grown plate (plate streaked from – 70°C glycerol stock about 12–16 h before use) of relevant culture and grown overnight (12–16 h) in a shaker-incubator (37°C, 200 rpm). Next day, 5–10 ml of medium with the appropriate antibiotic was seeded with overnight grown culture at 0·1% (v/v) level and grown up to an $OD_{600}$ of about 0·3. At this stage an appropriate volume of culture was induced with 1 mM (final concentration) IPTG and grown further for 5–6 h.

2.4b *Detection of gene expression: immuno-capturing and Western blot detection of rGH*: Cells from 10 ml of post-induced cultures were harvested and lysed with 1 ml of modified sample buffer [MSB: 150 mM Tris HCl, pH 6·8; 2% (w/v) SDS and 3 M urea]. Lysates were diluted 20-fold with the dilution buffer (0·2 M NaCl, 0·05% NP-40 and 50 mM Tris HCl, pH 7·2). To capture antigen, 10 ml of anti-$GH_{rabbit}$ antisera was added and the mixture was incubated at 4°C for 4–5 h with gentle rocking. To this 100 µl of a pre-swollen, washed slurry (approx. 1 : 1 suspension of beads) of protein-A-Sepharose® (Amersham Pharmacia, Uppsala, Sweden) was added and the rocking was continued for another 4–5 h at 4°C. The protein-A-Sepharose-antibody-antigen complex was then settled by

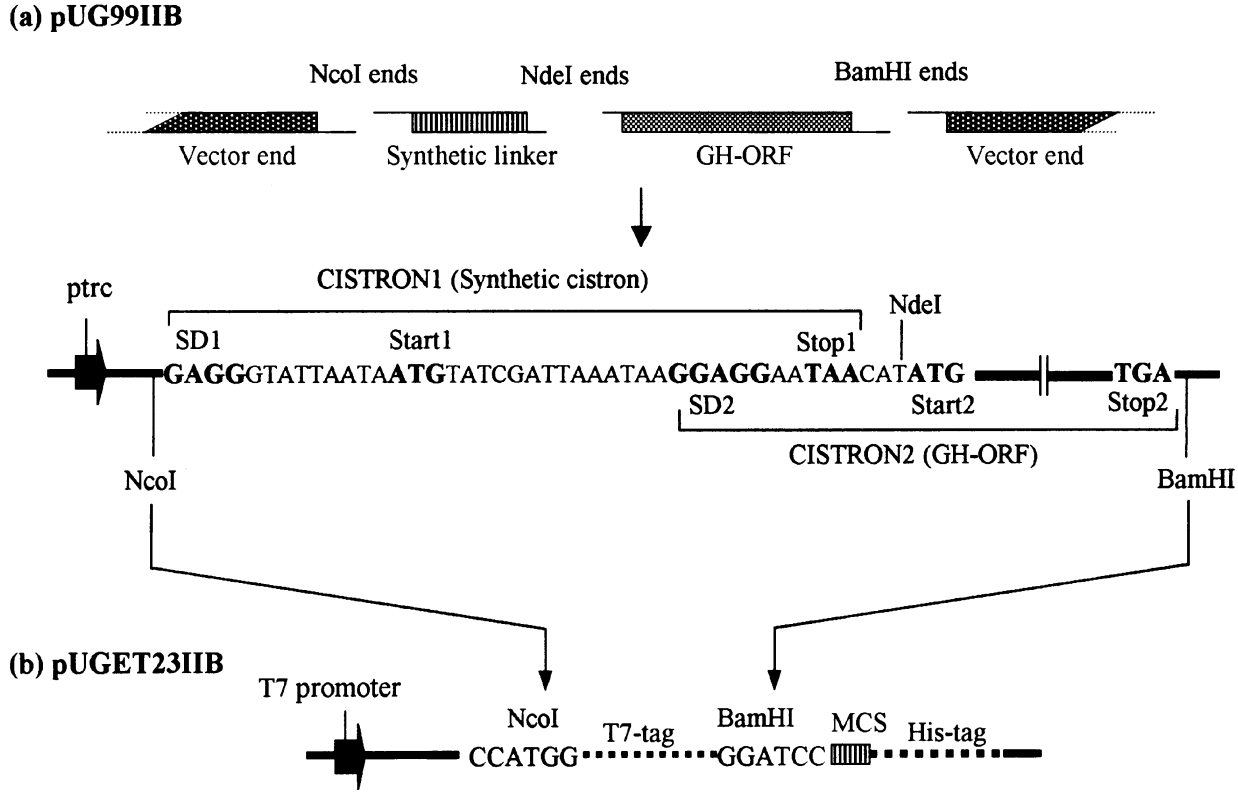


**Figure 1.** Synthetic two-cistronic GH-expression constructs under *trc* (pUG99IIB) and *T7* (pUGET23IIB) promoters, respectively, in pTrc99A and pET23d$^+$. Various translational signals viz. SD, ATG, stop codon, etc. are indicated (highlighted). (**a**) Development of pUG99IIB: In a three piece ligation reaction the GH-ORF on a *Nde*I-*Bam*HI fragment was ligated to *Nco*I-*Bam*HI digested vector pTrc99A through a synthetic linker carrying *Nco*I and NdeI sticky ends. (**b**) Construction of pUGET99IIB: The entire two-cistronic expression cassette was excised out from pUG99IIB and subcloned between *Nco*I-*Bam*HI sites in pET23d$^+$.

centrifugation at 5,000 rpm for 5 min. Beads were washed thrice with 1 ml of dilution buffer. Finally, bound protein was eluted in 100 µl SDS sample buffer [Tris HCl, pH 6·8 (50 mM), SDS (2%, w/v), glycerol (10%, v/v), *b*-mercaptoethanol (7·5%, v/v) and bromophenol blue (0·01%, w/v)] with boiling for 5 min.

For Western blotting, 25 ml of the above sample was resolved on 12·5% SDS-PAGE. The proteins from the gel were electrophoretically transferred onto NC-membrane (Towbin *et al* 1979). Transferred bands were probed using 1 : 1000 dilution of anti-$GH_{rabbit}$ antiserum and HRP-labelled anti-$immunoglobulin_{goat}$ antibody (1 : 5,000 dilution).

## 3. Results and discussion

Initially, when we cloned the native *GH* gene/s directly under the strong *trc* or the *T7* promoters, very low levels of GH expression were observed in both these cases. We then constructed a two-cistronic expression system for the *GH* gene/s (construct pUG99IIB) with the sequence

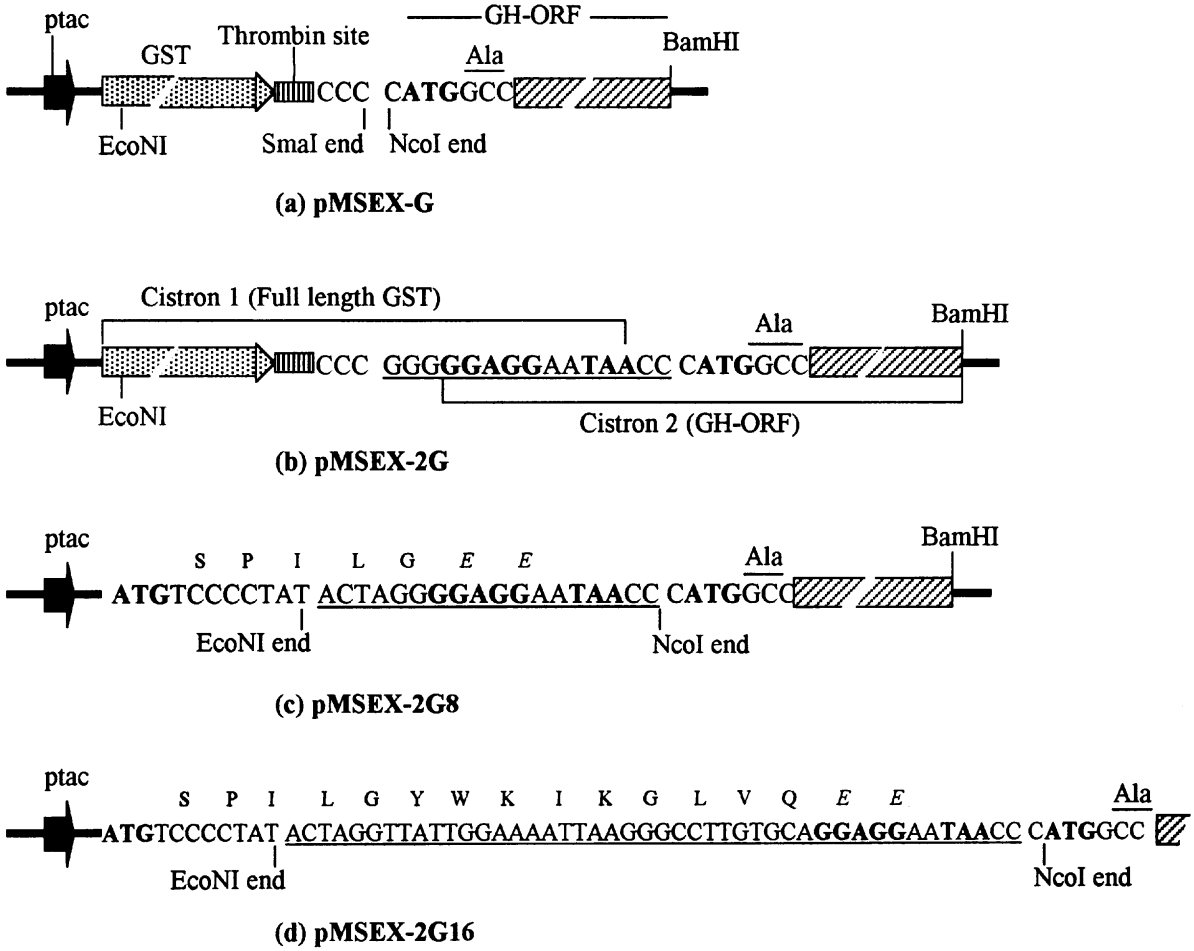


**Figure 2.** Development of various semi-synthetic *GST-GH* two-cistronic constructs. Only the sense strand sequences are shown. Various important sequence elements viz. promoter (ptac), thrombin cleavage site, restriction/cloning sites, translational signals (viz. SD, ATG, stop codon etc. shown in bold face), the first amino acid Ala (+ 1) of GH etc. are indicated. (**a**) Construction of GST-GH fusion construct pMSEX-G: *GH*-ORF present on a *Nco*I-*Bam*HI fragment with *Nco*I-end being blunted (using $T_4$ DNA polymerase-mediated fill in reaction) was cloned in-frame with *GST* at *Sma*I site in pGEX-KT. (**b**) Construction of pMSEX-2G: *GH*-ORF on a *Nco*I-*Sac*I fragment (with *Nco*I-end being polished by $T_4$ DNA polymerase) from pMSEX-G was subcloned through a linker (underlined) between *Sma*I and *Sac*I sites in pGEX-KG. (**c**) Construction of pMSEX-2G8: *Eco*NI and *Nco*I digested pMSEX-2G was religated through a synthetic linker molecule (underlined). (**d**) Same as above, except in terms of the length and sequence of the linker molecule incorporated. A bigger linker was used to include within the first cistron a larger portion of the native *GST* sequence. The amino acids encoded by the first cistrons (in panel **c** and **d**) are indicated (italicized amino acids are not natural to *GST* cistron).

of the first cistron very similar to that in pCZ145, as described (Schoner *et al* 1986). This construct resulted in high-level accumulation of GH (~ 30%) in the cytoplasm. In another construct (pMSEX-G), we fused the *GH* gene in-frame with *GST* (figure 2a) in pGEX-KT, which resulted in high level (20–25%) production of the GST-GH hybrid protein. To obtain native GH in an unfused state using the latter system, and also to know whether any highly expressible natural cistron can act as a first cistron in a two-cistron expression system as suggested by Schoner *et al* (1986), we constructed a *GST-GH* two-cistron expression system (pMSEX-2G) using full-length *GST* gene as the first cistron, in the expression vector pGEX-KG. The SD sequence, distance between SD and ATG of the downstream cistron, as well as the positioning of stop codon (stop 1) for the upstream cistron, was kept exactly similar to that of the earlier construct pUG99IIB which had given high level expression of GH. While this construct expressed GST well (to a level similar to that by the native vector pGEX-KG), no detectable GH expression was observed upon SDS-PAGE analysis. However, the fact that the GH expression did occur, albeit at

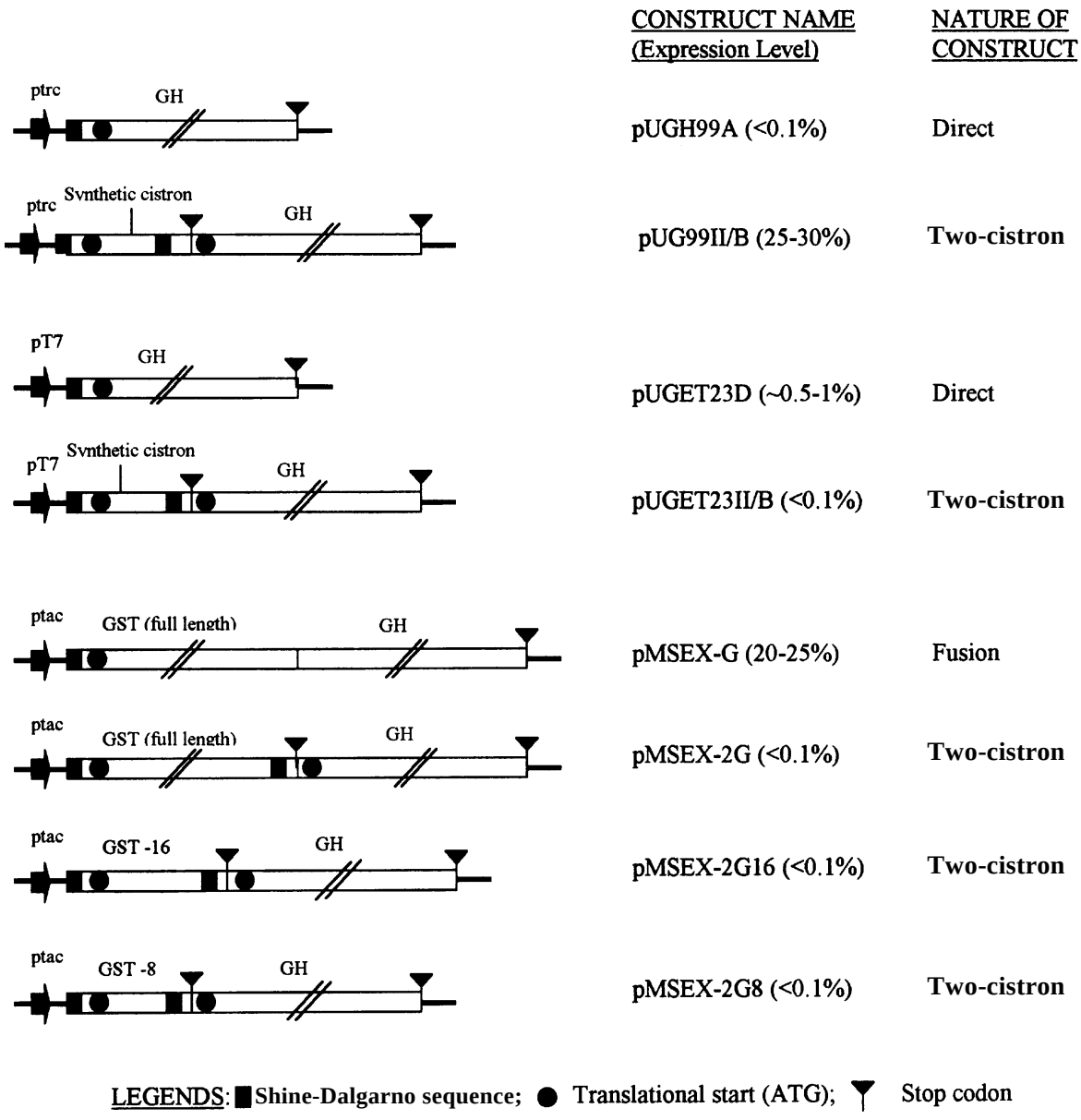


**Figure 3A.** Schematic representation of different GH-expression constructs. The levels of expression as indicated were estimated either by comparative Western blot technique or by densitometric scanning of Coomassie-stained polycarylamide gels.

much lower levels, was confirmed by Western blotting of immunoprecipitated samples, indicating the presence of a functional *GH* gene. We then constructed two semi-synthetic two-cistronic expression vectors viz, pMSEX-2G8 and pMSEX-2G16 using shorter lengths of *GST* gene as the first cistron to check whether, for some inexplicable reasons, a big cistron could not serve as an efficient first cistron. The length of the first cistron was either kept exactly the same (as in pMSEX-2G8) or was doubled (in pMSEX-2G16) in comparison to the first cistron of pUG99IIB, preserving as much *GST* sequence as permissible [since the incorporation of SD2 upstream of stop 1 resulted in a limited change of nucleotide sequence leading thereby to change in the last few amino acids of the portion of GST being used as the first cistron (cf. figure 2c, d)]. The SD sequence, the SD-ATG spacing of RBS2 and the relative location of the stop 1 in both these constructs was exactly reproduced from pUG99IIB. Both these constructs, however, failed to hyper-express GH; remarkably low levels of expression of GH were again confirmed by immunoprecipitation, thereby indicating the presence of a functional *GH* cistron in these constructs as well (see figure 3 for the approximate level of expression obtained using different GH-expression systems).

The mechanism proposed earlier (Schoner *et al* 1986, 1987) to explain the working of a two-cistronic system (see §1) fails to explain adequately the low level of GH expression observed with our *GST*-based two-cistronic constructs, especially in the case of pMSEX-2G, where

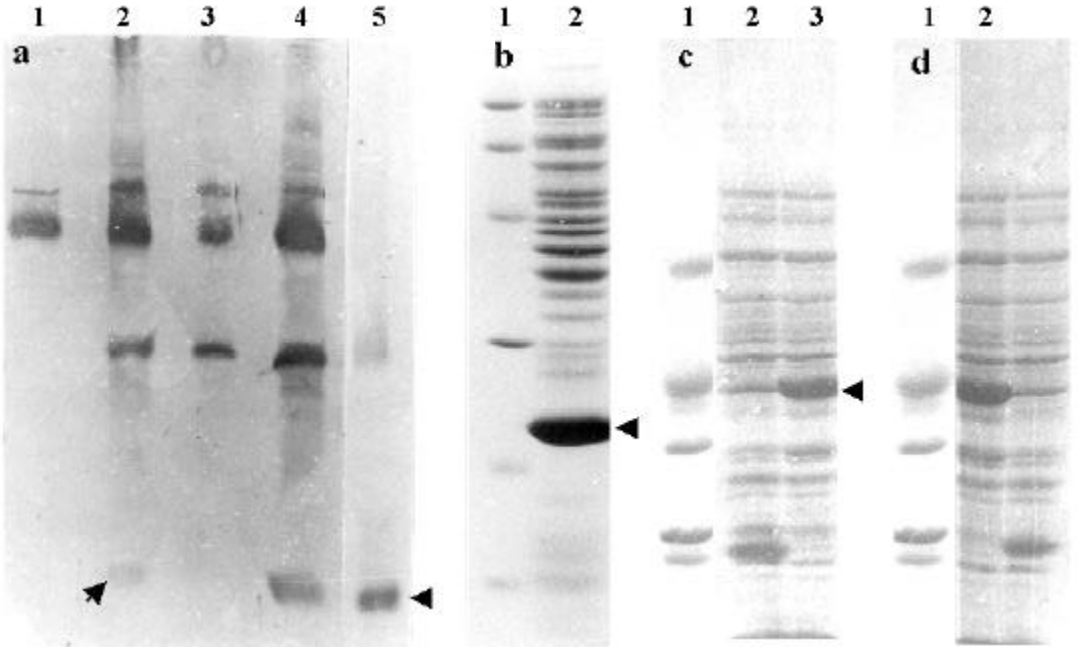


**Figure 3B.** Representative gels/blots showing expression of GH by some of the GH-expression plasmids. (**a**) Immuno-detection of expression of GH with pUGH99A. Ten ml of 5 h post-induced cultures of either *E. coli* JM105 carrying control plasmid pTrc99A or GH expression vector pUGH99A were harvested and lysed with 1 ml of modified sample buffer. After 20-fold dilution of the lysates, 10 µl of anti-$GH_{rabbit}$ antisera was added to capture GH in the lysates. The antigen-antibody complexes were then harvested with 100 µl of Protein-A Sepharose. Finally, the bound GH was released by boiling with 50 µl of SDS-sample buffer, a part (25 µl) of which was resolved through SDS-PAGE and processed for detection of GH through Western blotting (see § 2 for detail). Lane 1, 25 µl (1 : 1000 dilution) of anti-$GH_{rabbit}$ antisera directly run on gel; lane 2, control lyaste where 1 µg of std. pit. bGH was added prior to immunoprecipitation; lane 3, immunoprecipitated control lysate without added GH; lane 4, immunoprecipitated pUGH99A lysate, and lane 5, standard pituitary bGH (500 ng) directly run on SDS-PAGE. (**b**) GH hyper-expression in post-induced culture pUG99IIB as visualized on SDS-PAGE. Lane 1, indicates different standard protein markers with molecular weights in kDa (from top to bottom) being 97·4, 66·2, 42·7, 31, 21·5 and 14·4; lane 2, lysate of post-induced culture elaborating expression plasmid pUG99IIB showing very high level accumulation of GH. (**c**) SDS-PAGE analysis of expression of GST-GH fusion protein. Lane 1, indicates the different standard protein markers with molecular weights (from top to bottom) being 66, 45, 36, 29 and 24 kDa; lane 2, control lysate of *E. coli* XL1-Blue cells carrying plasmid vector pGEX-KT; lane 3, represents post-induced lysate prepared from clone carrying GST-GH fusion construct pMSEX-G. (**d**) SDS-PAGE analysis of expression of GH with GST (full-length)-GH two-cistronic construct pMSEX-2G. Lane 1, indicates the different standard protein markers as in panel **c**; lane 2, represents post-induced lysate of GST-GH fusion construct pMSEX-G; lane 3, represents post-induced lysate of clone carrying GST-GH two-cistronic gene expression construct pMSEX-2G.

initiation at RBS1 is highly efficient, as evidenced by high level GST expression (figure 3B, panel d). There is no apparent reason why the advantages offered by a terminating ribosome in improving the translational efficiency of the second cistron would not have existed in our situation. Therefore, the generality of the two-cistron approach, as has been claimed (Schoner *et al* 1987), seems questionable. Regarding designing of such a system and the choice of a first cistron, the authors opined that "in principle any sequence that allows for efficient ribosome binding and translation initiation should be suitable as a first cistron", and also suggested that

. . . two-cistron expression system can be derived directly from one-cistron expression systems that encode hybrid (or fusion) proteins . . . ". This simply entails the conversion of the mRNA encoding the fusion protein to a two-cistron mRNA. In general, such conversion requires only a few base changes that create a translational stop codon at the end of the *E. coli* sequence, a translational start codon for the eukaryotic gene and an SD sequence near the end of the *E. coli* sequence and 5′ to the beginning of the gene". However our studies, based on the use of different lengths of the N-terminal portion of a naturally highly expressible gene (i.e. *GST*) as the first cistron (e.g. as in pMSEX-2G, pMSEX-2G8 and pMSEX-2G16) clearly show that the translational efficiency of the *GH* gene (the second cistron) could not be improved by this strategy. We suspect the events at RBS2 that lead to hyper-expression of the second cistron to be more complex in nature than envisioned earlier (Schoner *et al* 1986, 1987). The successful initiation at RBS2 in the two-cistronic arrangement possibly depends upon a 'particular favourable sequence feature' around RBS2, which, in conjunction with the advantages offered

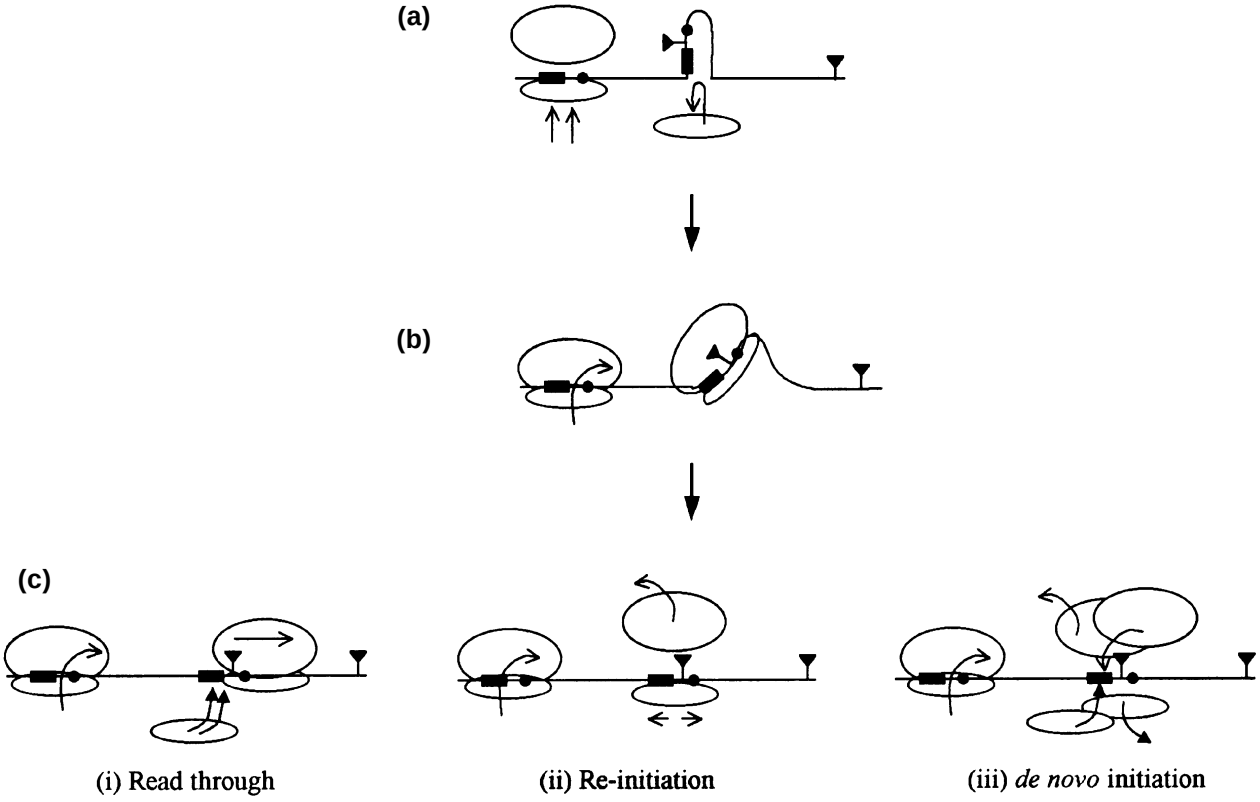


**Figure 4.** Schematic representation of various possible mechanisms of working of two-cistronic expression systems. (**a**) High frequency initiation of translation at structure-free RBS1 and failure of initiation by 30S at RBS2 due to the presence of strong secondary structure. (**b**) The 70S ribosomes when reach stop 1 physically disrupt the prevailing structure/s, setting RBS2 free for translational initiation by one or more of the following mechanisms (shown in panel **c**). (**c**) (i) In *readthrough* type of translation, the same 70S ribosome translating the first-cistron continues translating the second-cistron without being released from the mRNA. This allows many more *de novo* initiation by the cellular pool of ribosomes; (ii) in *re-initiation* mode, the 70S ribosome dissociates when it encounters stop 1, the 50S particle gets released, while the 30S remain attached to the mRNA, and scans back and fourth for neighbouring RBS, and reinitiates translation when it finds RBS2; (iii) in *de novo* initiation the ribosome translating upstream cistron, while reaching stop 1 disrupt the structure/s around RBS2 and dissociates and the translational initiation at RBS2 is carried out only by fresh/recycled ribosomal subunits.

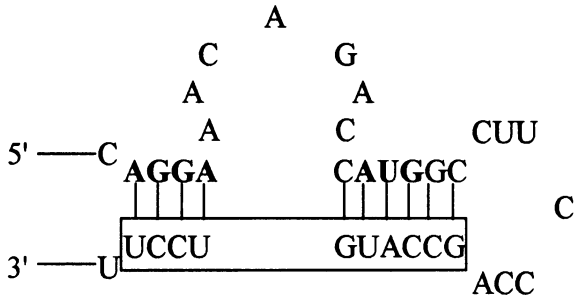


**Figure 5.** Internal negative control element (INCE): A stretch of approximately 10-base sequence (5′-GCCAUGUCCU-3′) within the first 20 bases of *GH*-coding sequence, bearing strong complementarity with RBS was identified using RNAFOLD programme. In this figure is shown the basepairing of INCE (boxed) with RBS in the construct pUGH99A, where native *GH*-ORF is placed directly under *trc* promoter (see figure 3).

by terminating ribosome/s, as proposed earlier, produces a potentiation effect.

There seem to be at least three overt possibilities by which the efficiency of translational initiation can improve at RBS2 in a two-cistron expression system (figure 4).

(i) A *read-through* type of translation, where the terminated 70S ribosome also initiates translation at RBS2 without being released from the mRNA.

(ii) A terminating ribosome dissociates when it encounters a stop codon in the first cistron, the 50S particle gets released to the free pool of cellular ribosomal subunits, while the 30S remains attached to the mRNA, and scans the downstream sequences for neighbouring RBS, and *reinitiates* translation when it finds one.

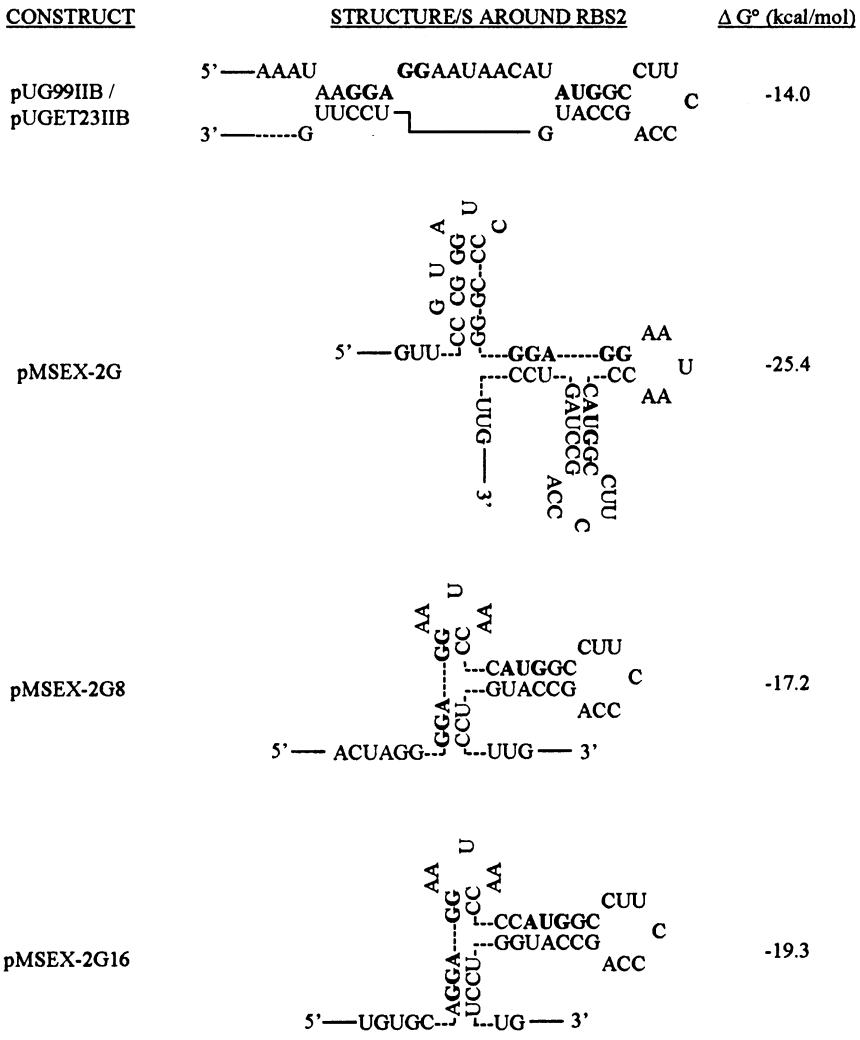


**Figure 6.** Secondary structure/s around RBS2 in various *GH* two-cistronic expression constructs. The prevailing structures along with the respective ΔG° values were computed using computer software RNAFOLD. The SDs and ATGs are shown in bold face.

(iii) The higher frequency of *de novo* initiation by fresh/recycled ribosomal subunits favoured due to "opening up" of RBS2 by the terminating 70S ribosomes and/or due to the increased local concentration of ribosomal subunits.

It is well established that ribosomes are efficiently released at the termination codon in the presence of ribosome releasing factor (RRF) and it is unlikely that in cells in which RRF is functioning, the same ribosome continues to translate one cistron to the next (Nakamura *et al* 1996; Kisselev and Buckingham 2000). Based on recent advances on the understanding of translation termination and the role of RRF in disassembling the termination complex (Janosi *et al* 1998; Kaji *et al* 1998), one can presume that there is no 'read-through' type of mechanism existing in this situation, and even if it exists, the contribution of such a mechanism is probably negligible. Otherwise also, a read-through type of mechanism would have ensured high-level GH expression, at least in pMSEX-2G, where GST expression is normal indicating efficient ribosome entry through RBS1. To explain our results with the *GST*-based two-cistronic expression constructs, we assume that the translating ribosome (entering at RBS1), upon encountering a stop codon (an 'ochre' codon in our studies) in the absence of a suppresser tRNA and in the presence of functional RRF and IF3, invariably dissociates, although the 30S subunit may still remain bound to the mRNA. This 30S particle can scan the message by lateral diffusion, as has been suggested (Audin and van Duin 1990), in search of a neighbouring initiation sequence and thereby reinitiate translation. In earlier studies (Danchin and Ullmann 1980; Brot *et al* 1981) it had been proposed that ribosomes can glide on the intercistronic region of mRNA in search of the next initiation codon without polypeptide synthesis ("phaseless" and "sterile" travel). Once such a ribosome successfully reinitiates translation, it allows many more fresh initiation events to take place by cellular ribosomal subunits (de Smit and van Duin 1990), and it has also been shown that binding of 30S subunit shifts the equilibrium in the direction of unfolding of mRNA. The success of reinitiation depends upon the stability of the structure in and around the downstream RBS. Since such a scanning process by the 30S ribosomal subunit is considered not to be an energy requiring process (Audin and van Duin 1990), any structure downstream would stall the scanning ribosome, thereby preventing reinitiation. Remarkably, there exists a small stretch of 10-base sequence immediately downstream of the ATG codon in the *GH* gene/s, that shows a strong duplex formation potential with the RBS sequence in the

| | Start | Ser | Pro | Ile | Leu | Gly | Tyr | Trp | Lys | Ile | Lys | Gly | Leu | Val | Gln | Pro | Thr | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | CUG | | | | | | | | | | | | | |
| | | | | | UUG | | | | | | | | | | | | | |
| | | | | | UUA | | | | | | | | CUG | | | | | |
| | | AGC | | AUC | CUU | | | | | | AUC | | UUG | | | | ACC | |
| | | UCU | CCG | AUU | CUC | GGC | UAC | | AUC | | AAA | | UUA | | CAG | | ACG | |
| **a) pMSEX-2G:** | **AUG** | **UCC** | **CCU** | **AUA** | **CUA** | **GGU** | **UAU** | **UGG** | **AAA** | **AUU** | **AAG** | **GGC** | **CUU** | **GUG** | **CAA** | **CCC** | **ACU** | **----** |
| | | UCG | CCA | | | GGG | | | AAG | AUA | | GGU | CUC | GUU | | CCA | ACA | |
| | | AGU | CCC | | | GGA | | | | | | GGG | CUA | GUC | | CCU | | |
| | | UCA | | | | | | | | | | GGA | | GUA | | CCG | | |
| | | | | | | | | | | | | | | | | ***Glu*** | ***Glu*** | **Stop** |
| | | | | | | | | | | | | | | | | | GAG | |
| **b) pMSEX-2G16:** | **AUG** | **UCC** | **CCU** | **AUA** | **CUA** | **GGU** | **UAU** | **UGG** | **AAA** | **AUU** | **AAG** | **GGC** | **CUU** | **GUG** | **CA*G*** | ***GAG*** | ***GAA*** | **UAA** |
| | | | | | | | | | | | | | | | CAA | GAA | | |
| | | | | | | | ***Glu*** | ***Glu*** | **Stop** | | | | | | | | | |
| | | | | | | GGC | | | | | | | | | | | | |
| | | | | | | GGU | | GAG | | | | | | | | | | |
| **c) pMSEX-2G8:** | **AUG** | **UCC** | **CCU** | **AUA** | **CUA** | **GG*G*** | ***GAG*** | ***GAA*** | **UAA** | | | | | | | | | |
| | | | | | | | GAA | | | | | | | | | | | |

**Figure 7.** The codon usage pattern in various first cistrons in *GST*-based two-cistronic constructs. The codon used (to specify the 5′-region of the first cistron in case of pMSEX-2G, and the entire first cistron of pMSEX-2G16 and pMSEX-2G8) are shown in bold face. Various codons utilized for each amino acid are also shown against their respective positions, arranged in order of preference of usage in *E. coli* (Wada *et al* 1991) with the most preferred codon being placed at the top, and the least at the bottom. Amino acids encoded by these cistrons or part of the cistron are shown at the top. Subsequently, changed residues (italicized) which are not native to the GST are also indicated.

*GH* message emanating from the *GST-GH* constructs. Therefore, it is likely that there prevails an open competition for RBS sequences between the 30S ribosome and this "internal negative control element" [abbreviated as INCE (figure 5)], which is temporarily pushed out by terminating 70S ribosomes, leaving RBS2 transiently free of structure. It follows that if the affinity of the 30S subunit for the mRNA ($\Delta G°_{30S}$) is more than the affinity of INCE for RBS ($\Delta G°_{structure}$), the 30S subunit would succeed in reinitiating translation, failing which the RBS2 would be "recaptured" by INCE. Here, the $\Delta G°_{30S}$ and $\Delta G°_{structure}$ are the free energy changes due to pairing of RBS with 16S rRNA and RBS with INCE respectively.

In keeping with the above model, the possible secondary structure potential of the sequence around RBS2 in the two-cistronic constructs were analysed using Zuker's RNAFOLD programme (Jackobson *et al* 1984). It was found that INCE alone possesses higher affinity for RBS (figure 6) than the average 30S subunit's affinity for mRNA in all our constructs, including pUG99IIB, which hyper-express GH (the average $\Delta G°_{30S}$ is around – 10 kcal/mol as determined *in vitro* for many efficient RBS (Calogera *et al* 1988; Schurr *et al* 1993; Osada *et al* 1999). However, although the higher affinitiy of INCE for RBS2 explains well the apparently low probability of reinitiation in constructs pMSEX-2G, pMSEX-2G8 and pMSEX-2G16 (showing $\Delta G°_{structure}$ values of – 25·4,

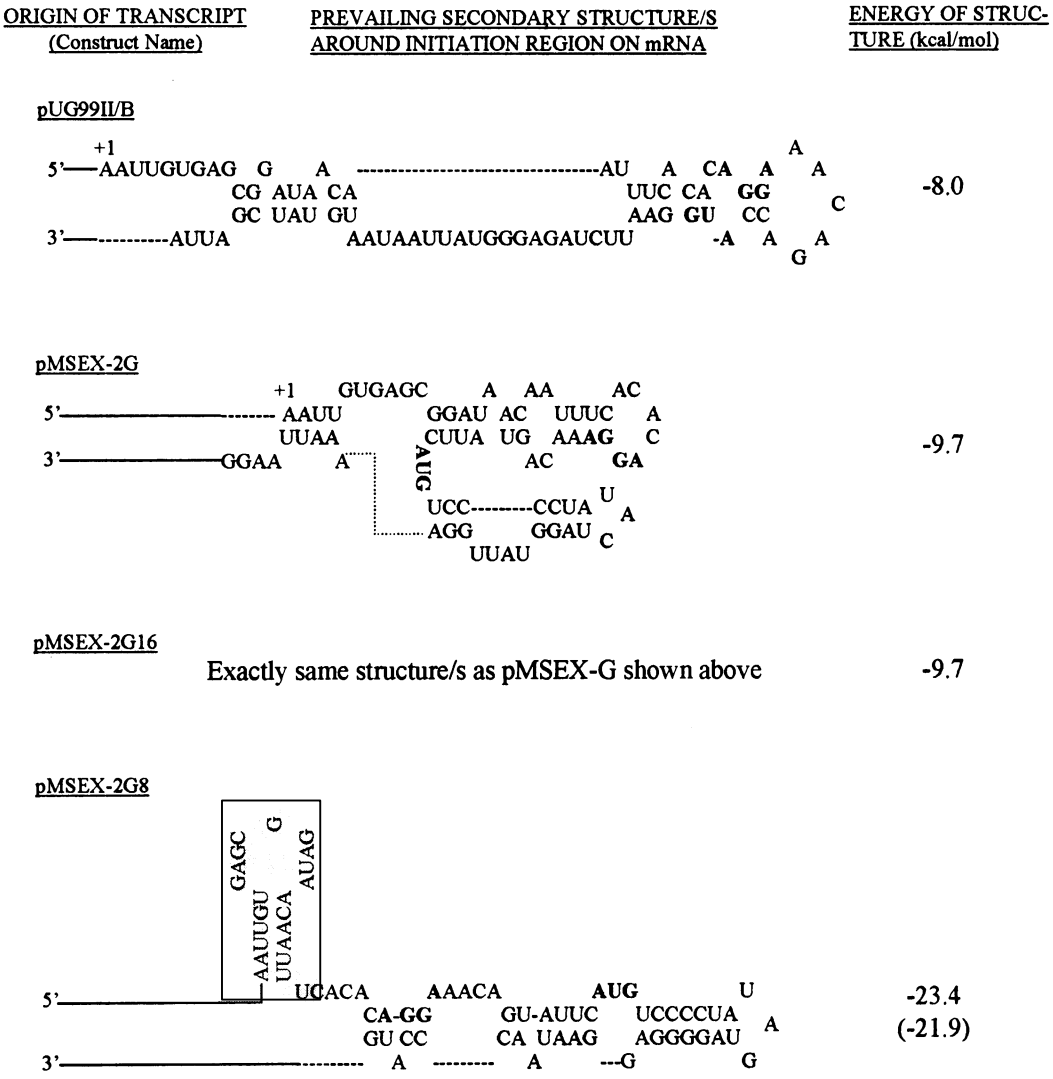


**Figure 8.** The mRNA secondary structures (of the first 75 nucleotides) as predicted by Zuker's RNAFOLD programme. The SDs and start codons are shown in bold face. The value within parenthesis is the recalculated energy of structure obtained upon reanalysing the sequence (21st to 75th nucleotide) after eliminating the region forming the hairpin loop (shaded).

– 17·2, – 19·3 kcal/mol respectively), it partially fails to explain the high expressibility of the *GH* gene in pUG99IIB as it has slightly higher $\Delta G°_{structure}$ (– 14 kcal/mol) than the average $\Delta G°_{30S}$ (– 10 kcal/mol).

To explain this, we propose a scenario wherein reinitiation at RBS2 in pUG99IIB takes place when INCE either has not yet been transcribed or is still in association with RNA polymerase and/or the template strand of DNA. It is known that approximately the last 25 ribonucleotides added to a growing mRNA chain are complexed with DNA and/or enzyme at any moment (Lewin 1999). In a coupled transcription-translation mode, ribosomes could enter at structure-free RBS1 as soon as it is available, chase RNA polymerase with almost equal speed [the *E. coli* RNA polymerase adds ~ 40 ribonucleotides/s, and the average rate of translation is about 15 amino acids/s (Lewin 1999)], terminate at stop 1 and reinitiate at RBS2 before INCE appears in the fray (figure 9a). It is quite possible that reinitiation takes very little time because of the close proximity of the 30S subunit to RBS2 (at least, less time is required than a completely fresh initiation).

With the high-level expression of GST in pMSEX-2G, it became quite evident that the ribosomes indeed enter through RBS1, with high frequency, but possibly even the first ribosome reaches stop 1 lagging behind the RNA polymerase (figure 9b). This can happen because, although RNA polymerase migrates almost at a constant average speed, the ribosomal movement is highly variable depending upon the codons it has to decode (Pedersen 1984; Sorensen *et al* 1989; Solomovici *et al* 1997; Lesnik *et al* 2000). It is known that the time taken by a ribosome

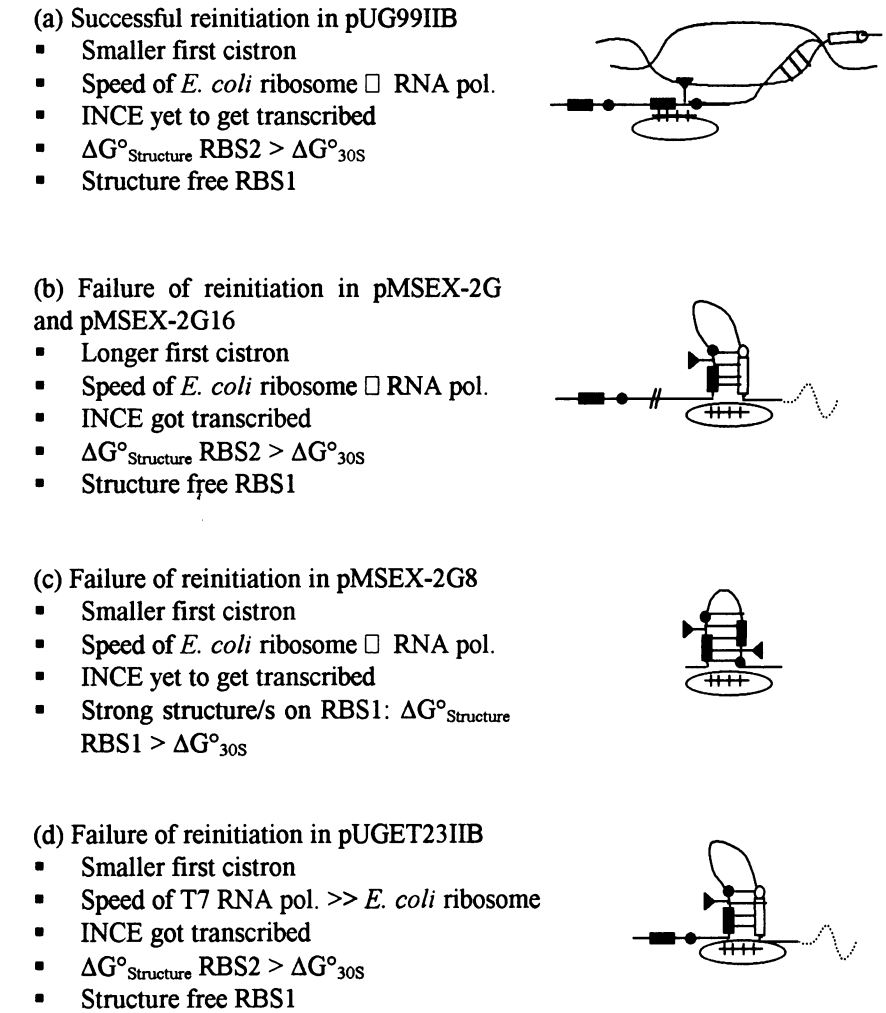


**Figure 9.** Schematic representation of possible scenarios prevailing with different GH two-cistronic expression systems leading to successful/unsuccessful reinitiation events at RBS2.

to read can be as short as 0·6 s for a 'rare codon' and as long as 2 s for a 'rare tRNA codon' subset of the rare codons (Pedersen 1984). [It was found that *GST* indeed utilizes many rare codons to specify the first few amino acids of the polypeptide (see figure 7)].

Although ribosome entry at RBS1 in pMSEX-2G16 has not been established, the secondary structure analysis of the first 75 nucleotides including RBS1, show similar $\Delta G°_{structure}$ with pMSEX-2G (– 9·7 kcal) and close to that of pUG99IIB (– 8·0 kcal) (see figure 8). Therefore, it can be assumed that the initiation takes place at RBS1, but possibly because of the slower ribosome movement and a longer first cistron (cistron 1 here is longer by 20–25 bases than Schoner *et al*'s synthetic cistron), the ribosomes lag behind RNA polymerase. Thus, by the time the first ribosome reaches stop 1 the INCE would get transcribed, and thus rendered free to compete with the 30S subunit for pairing with RBS2 (figure 9b).

The poor expression observed in the case of pMSEX-2G8, where the first cistron length was kept exactly similar to Schoner *et al*'s cistron, possibly has resulted due to two reasons; firstly, the secondary structure analysis of first 75 bases showed a drastic reduction of $\Delta G°_{structure}$ (– 23·4 kcal). Although it utilizes the GST RBS ($\Delta G°_{structure}$ = – 9·7), due to the change of the sequence context (as a result of the presence of the G-C rich 5′-end of *GH* gene), strong secondary structure/s prevails in the region of the RBS1 itself (see figure 8), which likely affects the ribosome entry at RBS1 in the first place. Here the sequence on and around RBS2 were actually found to pair strongly with RBS1 ($\Delta G°_{structure}$ = – 21·9 kcal/mol), acting as a negative control element for the RBS1 (figure 9c). Thus, apart from the INCE type of sequence emanating from the second cistron, the sequence of first cistron, RBS2 and their potential to form strong structure with RBS1, if not checked carefully might also become the source of major problem.

To test whether pUG99IIB expresses GH as per our proposed mechanism, we needed to decouple the transcription and translation to show that in the presence of free INCE, ribosomes cannot reinitiate optimally. An *in vitro* transcription followed by *in vitro* translation of the mRNA utilizing *E. coli* transcription and translation machineries would have perhaps best served this purpose. Nevertheless, in order to explore this aspect, we adopted an alternate strategy which can functionally decouple or, at least widen the time gap, between the process of transcription and translation. The entire two-cistronic cassette (RBS1 to stop 2) was subcloned under *T7* promoter in pET23d$^+$ (figure 1b). It is known that transcription by the T7 RNA polymerase is much faster [which polymerizes nearly 200 nucleotides/s (Lewin 1999)] than the rate of translation by the fastest *E. coli* ribosome (up to about 20 amino acids, i.e. 60 N/s). Thus, the T7 polymerase will have transcribed almost half the *GH* gene by the time the first ribosome reaches stop 1 (figure 9d). Entirely in keeping with our reasoning, this construct (pUGET23II/B) failed to hyper-express GH although it expressed GH at low levels, as confirmed by Western blotting.

It appears that the designing of two-cistron expression systems is not a straightforward task, and it would probably be more a matter of chance if a two-cistron system designed for one gene turned out to be equally effective for another. In other words, for optimising the expression of a given gene using the two-cistron approach, the first cistron sequence has to be tailor-made such that it bears minimum complementarity with RBS1, RBS2 and the 5′-end of the downstream gene to be expressed. The length and codons chosen to specify the first cistron could be the most important and probably the limiting factor in a situation where the gene has an INCE type of sequence with high affinity (higher than the $\Delta G°_{30S}$) for sequences in close proximity to RBS2. We conclude that a natural cistron, even though allowing efficient translational initiation, may not be a suitable alternative to a synthetic cistron carefully designed with these considerations in mind, because a highly efficient RBS may turn into a poorly translatable one as a result of a changed sequence context.

## Acknowledgements

The work was funded by the Council of Scientific and Industrial Research, New Delhi. The standard pituitary bGH and anti-bGH antisera were generous gifts from Dr A F Parlow, Pituitary Hormone and Antisera Centre, Harbor-UCLA Medical Center, Torrance, California, USA.

## References

Amann E 1988 Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*; *Gene* **69** 301–315

Audin H R and van Duin J 1990 Scanning model for translational reinitiation in Eubacteria; *J. Mol. Biol.* **213** 811–818

Brot N, Caldwell P and Weissbach H 1981 Regulation of synthesis of *Escherichia coli* ribosomal protein L1 and L11; *Arch. Biochem. Biophys.* **206** 51–53

Calogera R A, Pow C L, Canonaca M A and Gualerzi C O 1988 Selection of mRNA translation initiation region by *Escherichia coli* ribosomes; *Proc. Natl. Acad. Sci. USA* **85** 6427–6431

Danchin A and Ullmann A 1980 The coordinate expression of polycistronic operons in bacteria; *Trends. Biochem. Sci.* **2** 51–52

Das A and Yanofsky C 1984 A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair; *Nucleic Acids Res.* **12** 4757–4768

de Smit M H and van Duin J 1990 Secondary structure of ribosome binding site determines translational efficiency: a quantitative analysis; *Proc. Natl. Acad. Sci. USA* **87** 7668–7672

de Smit M H and van Duin J 1994 Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data; *J. Mol. Biol.* **244** 144–150

Guan K and Dixon J E 1991 High-level expression of prochymosin in *Escherichia coli*: effect of the secondary structure of the ribosome binding site; *Anal. Biochem.* **192** 262–267

Hakes D J and Dixon J E 1992 New vectors for high level expression of recombinant proteins in bacteria; *Anal. Biochem.* **202** 293–298

Jacobson A B, Good L, Simonetti J and Zuker M 1984 Some simple computational methods to improve the folding of large RNAs; *Nucleic Acids Res.* **12** 45–52

Janosi L, Mottagui-Tabar S, Isaksson L A, Sekine Y, Ohtsubo E, Zhang S, Goon S, Nelken S, Shuder M and Kaji A 1998 Evidence for *in vivo* ribosome recycling, the fourth step in protein biosynthesis; *EMBO J.* **17** 1141–1151

Kaji A, Teyssier E and Hirokawa G 1998 Disassembly of post-termination complex and reduction of translational error by ribosome recycling factor (RRF) – a possible new target for antibacterial agents; *Biochem. Biophys. Res. Comun.* **250** 1–4

Kisselev L L and Buckingham R H 2000 Translational termination comes of age; *Trends Biochem. Sci.* **25** 561–566

Lesnik T, Solomovici J and Reiss C 2000 Ribosome traffic in *E. coli* and regulation of gene expression; *J. Theor. Biol.* **202** 175–185

Lewin B (ed.) 1999 *Genes VII* (Oxford: Oxford University Press)

Linder P and Tuite M F 1999 The versatility of RNA structure and function; *Trends Genet.* **15** 302–303

Nakamura Y, Ito K and Isaksson A L 1996 Emerging understanding of translational termination; *Cell* **87** 147–150

Osada Y, Saito R and Tomita M 1999 Analysis of base pairing potentials between 16S rRNA and 5′UTR for translation initiation in various prokaryotes; *Bioinformatics* **15** 578–581

Pedersen S 1984 *Escherichia coli* ribosome translate *in vivo* with variable rate; *EMBO J.* **3** 2895–2898

Sambrook J and Russell D W (eds) 2001 *Molecular cloning*: *a laboratory manual* (New York: Cold Spring Harbor Laboratory Press)

Sanger F, Niklen S and Coulson A R 1977 DNA sequencing with chain terminating inhibitors; *Proc. Natl. Acad. Sci. USA* **74** 5463–5467

Schoner B E, Belagaje R M and Schoner R G 1986 Translation of a synthetic two-cistron mRNA in *Escherichia coli*; *Proc. Natl. Acad. Sci. USA* **83** 8506–8510

Schoner B E, Belagaje R M and Schoner R G 1987 Expression of eukaryotic genes in *Escherichia coli* with a synthetic two-cistron system; *Methods Enzymol.* **153** 401–416

Schurr T, Nadir E and Margalit H 1993 Identification and characterization of *E. coli* ribosome binding sites by free energy computation; *Nucleic Acids Res.* **21** 4019–4023

Solomovici J, Lesnik T and Reiss C 1997 Does *Escherichia coli* optimize the economics of translation process; *J. Theor. Biol.* **185** 511–521

Sorensen M A, Kurland C G and Pedersen S 1989 Codon usage determines translation rate; *J. Mol. Biol.* **207** 365–377

Studier F W, Rosenberg A H, Dunn J J and Dubendorf J W 1990 Use of T7 polymerase to direct expression of cloned genes; *Methods Enzymol.* **185** 60–89

Towbin H, Staehelin T and Gordon J 1979 Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications; *Proc. Natl. Acad. Sci. USA* **76** 4350–4354

van Meerten D, Girard G and van Duin J 2001 Translational control by delayed RNA folding: identification of kinetic trap; *RNA* **7** 483–494

Wada K, Wada Y, Doi H, Ishibashi F, Gojobori T and Ikemura T 1991 Codon usage tabulated from GenBank genetic sequence data; *Nucleic Acids Res.* **19** 1981–1986

Wang G, Liu N and Yang K 1995 High level expression of prochymosin in *Escherichia coli*: effect of secondary structure of ribosome binding site; *Protein Exp. Purif.* **6** 284–290

Wikstrom P M, Lind L K, Berg D E and Bjork G R 1992 Importance of mRNA folding and start codon accessibility in the expression of genes in a ribosomal protein operon of *Escherichia coli*; *J. Mol. Biol.* **224** 949–966

*MS received 26 December 2001; accepted 28 February 2002*

Corresponding editor: SEYED E HASNAIN