MICROBIOLOGICAL REVIEWS, Sept. 1996, p. 512–538
0146-0749/96/$04.00+0
Copyright © 1996, American Society for Microbiology
Vol. 60, No. 3

Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*†

SAVVAS C. MAKRIDES*

Department of Molecular Biology, T Cell Sciences, Inc., Needham, Massachusetts 02194

INTRODUCTION

The choice of an expression system for the high-level production of recombinant proteins depends on many factors. These include cell growth characteristics, expression levels, intracellular and extracellular expression, posttranslational modifications, and biological activity of the protein of interest, as well as regulatory issues in the production of therapeutic proteins (191, 254). In addition, the selection of a particular expression system requires a cost breakdown in terms of process, design, and other economic considerations. The relative merits of bacterial, yeast, insect, and mammalian expression systems have been examined in detail in an excellent review by Marino (362). In addition, Datar et al. (121) have analyzed the economic issues associated with protein production in bacterial and mammalian cells.

The many advantages of *Escherichia coli* have ensured that it remains a valuable organism for the high-level production of recombinant proteins (177a, 197, 254, 362, 406, 426, 510). However, in spite of the extensive knowledge on the genetics and molecular biology of *E. coli*, not every gene can be expressed efficiently in this organism. This may be due to the unique and subtle structural features of the gene sequence, the stability and translational efficiency of mRNA, the ease of protein folding, degradation of the protein by host cell proteases, major differences in codon usage between the foreign gene and native *E. coli*, and the potential toxicity of the protein to the host. Fortunately, some empirical "rules" that can guide the design of expression systems and limit the unpredictability of this operation in *E. coli* have emerged. The major drawbacks of *E. coli* as an expression system include the inability to perform many of the posttranslational modifications found in eukaryotic proteins, the lack of a secretion mechanism for the efficient release of protein into the culture medium, and the limited ability to facilitate extensive disulfide bond formation. On the other hand, many eukaryotic proteins retain their full biological activity in a nonglycosylated form and therefore can be produced in *E. coli* (see, e.g., references 170, 342, and 486). In addition, some progress has been made in the areas of extracellular secretion and disulfide bond formation, and these will be examined.

The objectives of this review are to integrate the extensive published literature on gene expression in *E. coli*, to focus on expression systems and experimental approaches useful for the overproduction of proteins, and to review recent progress in this field. Areas that have been covered in detail in recent reviews are included in abbreviated form in order to present their key conclusions and to serve as a source for further reading. As a matter of definition, the terms "periplasmic expression" and "extracellular secretion" will be used to refer to the targeting of protein to the periplasm and the culture medium, respectively, to avoid confusion.

* Mailing address: Department of Molecular Biology, T Cell Sciences, Inc., 119 4th Ave., Needham, MA 02194.

† This review is dedicated to the memory of William John Steele, an inspired scientist, a great man, mentor, and friend, who died on 8 December 1995. The world is a better place because of him.

Downloaded from http://mmbr.asm.org/ on October 5, 2015 by guest

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

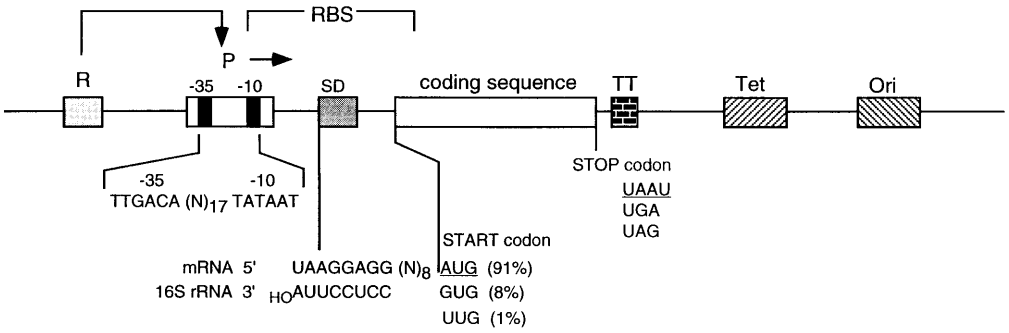

FIG. 1. Schematic presentation of the salient features and sequence elements of a prokaryotic expression vector. Shown as an example is the hybrid *tac* promoter (P) consisting of the −35 and −10 sequences, which are separated by a 17-base spacer. The arrow indicates the direction of transcription. The RBS consists of the SD sequence followed by an A+T-rich translational spacer that has an optimal length of approximately 8 bases. The SD sequence interacts with the 3′ end of the 16S rRNA during translational initiation, as shown. The three start codons are shown, along with the frequency of their usage in *E. coli*. Among the three stop codons, UAA followed by U is the most efficient translational termination sequence in *E. coli*. The repressor is encoded by a regulatory gene (R), which may be present on the vector itself or may be integrated in the host chromosome, and it modulates the activity of the promoter. The transcription terminator (TT) serves to stabilize the mRNA and the vector, as explained in the text. In addition, an antibiotic resistance gene, e.g., for tetracycline, facilitates phenotypic selection of the vector, and the origin of replication (Ori) determines the vector copy number. The various features are not drawn to scale.

CONFIGURATION OF EFFICIENT EXPRESSION VECTORS

The construction of an expression plasmid requires several elements whose configuration must be carefully considered to ensure the highest levels of protein synthesis (22, 64, 120, 142, 355, 538, 612). The essential architecture of an *E. coli* expression vector is shown in Fig. 1. The promoter is positioned approximately 10 to 100 bp upstream of the ribosome-binding site (RBS) and is under the control of a regulatory gene, which may be present on the vector itself or integrated in the host chromosome. Promoters of *E. coli* consist of a hexanucleotide sequence located approximately 35 bp upstream of the transcription initiation base (−35 region) separated by a short spacer from another hexanucleotide sequence (−10 region) (174, 232, 236, 344, 465). There are many promoters available for gene expression in *E. coli*, including those derived from gram-positive bacteria and bacteriophages (Table 1). A useful promoter exhibits several desirable features: it is strong, it has a low basal expression level (i.e., it is tightly regulated), it is easily transferable to other *E. coli* strains to facilitate testing of a large number of strains for protein yields, and its induction is simple and cost-effective (612).

Downstream of the promoter is the RBS, which spans a region of approximately 54 nucleotides bound by positions −35 (±2) and +19 to +22 of the mRNA coding sequence (269). The Shine-Dalgarno (SD) site (514, 515) interacts with the 3′ end of 16S rRNA during translation initiation (133, 532). The distance between the SD site and the start codon ranges from 5 to 13 bases (93), and the sequence of this region should eliminate the potential of secondary-structure formation in the mRNA transcript, which can reduce the efficiency of translation initiation (198, 229). Both 5′ and 3′ regions of the RBS exhibit a bias toward a high adenine content (140, 499, 502).

The transcription terminator is located downstream of the coding sequence and serves both as a signal to terminate transcription (465) and as a protective element composed of stem-loop structures, protecting the mRNA from exonucleolytic degradation and extending the mRNA half-life (35, 37, 147, 227, 249, 597).

In addition to the above elements that have a direct impact on the efficiency of gene expression, vectors contain a gene that confers antibiotic resistance on the host to aid in plasmid selection and propagation. Ampicillin is commonly used for this purpose; however, for the production of human therapeutic proteins, other antibiotic resistance markers are preferable to avoid the potential of human allergic reactions (42). Finally, the copy number of plasmids is determined by the origin of replication. In specific cases, the use of runaway replicons results in massive amplification of plasmid copy number concomitant with higher yields of plasmid-encoded protein (387, 415). In other cases, however, there appeared to be no advantage in using higher-copy-number plasmids over pBR322-based vectors (612). Furthermore, Vasquez et al. (572) reported that increasing the copy number of the plasmid decreased the production of trypsin in *E. coli* and Minas and Bailey (379) found that the presence of strong promoters on high-copy-number plasmids severely impaired cell viability.

TRANSCRIPTIONAL REGULATION

Promoters

A promoter for use in *E. coli* (Table 1) should have certain characteristics to render it suitable for high-level protein synthesis (207, 612). First, it must be strong, resulting in the accumulation of protein making up 10 to 30% or more of the total cellular protein.

Second, it should exhibit a minimal level of basal transcriptional activity. Large-scale gene expression preferably employs cell growth to high density and minimal promoter activity, followed by induction or derepression of the promoter. The tight regulation of a promoter is essential for the synthesis of proteins which may be detrimental to the host cell (see, e.g., references 68, 137, 544, 563, and 599). For example, the toxic rotavirus VP7 protein effectively kills cells and must be produced under tightly regulated conditions (592). However, in some cases, promoter stringency is inconsequential, because even the smallest amount of gene product drastically curtails bacterial survival because of its severe toxicity (615). For example, molecules that inactivate ribosomes or destroy the membrane potential would be lethal. Toxicity to the host is not restricted to foreign genes but may also result from the overexpression of certain native genes, such as the *traT* gene, which encodes an outer membrane lipoprotein (423), the *Eco*RI restriction endonuclease in the absence of the corresponding protective *Eco*RI modification methylase (423), and the *lon* gene (558). Furthermore, incompletely repressed expression systems can cause plasmid instability, a decrease in cell growth rate, and loss of recombinant protein production (40, 98, 374).

Downloaded from http://mmbr.asm.org/ on October 5, 2015 by guest

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

TABLE 1. Promoters used for the high-level expression of genes in *E. coli*

Promoter (source)	Regulation	Induction	Reference(s)
lac (*E. coli*)	*lacI*, *lacI*^{q}	IPTG	17, 18, 221, 460, 610
	lacI(Ts),[a] *lacI*^{q}(Ts)[a]	Thermal	234
	lacI(Ts)[b]	Thermal	604
trp (*E. coli*)		Trp starvation, indole acrylic acid	365, 470, 549, 612
lpp (*E. coli*)		IPTG, lactose[c]	128a, 142, 185, 275, 401
phoA (*E. coli*)	*phoB* (positive), *phoR* (negative)	Phosphate starvation	84, 274, 291, 306, 382, 562
recA (*E. coli*)	*lexA*	Nalidixic acid	145, 260, 428, 516
araBAD (*E. coli*)	araC	L-Arabinose	554
proU (*E. coli*)		Osmolarity	247
cst-1 (*E. coli*)		Glucose starvation	564
tetA (*E. coli*)		Tetracycline	125, 523
cadA (*E. coli*)	*cadR*	pH	102, 480, 561
nar (*E. coli*)	*fnr* (FNR, NARL)	Anaerobic conditions, nitrate ion	335
tac, hybrid (*E. coli*)	*lacI*, *lacI*^{q}	IPTG	7, 123, 471
	lacI[d]	Thermal	603
trc, hybrid (*E. coli*)	*lacI*, *lacI*^{q}	IPTG	65
	lacI(Ts),[a] *lacI*^{q}(Ts)[a]	Thermal	4, 9
lpp-lac, hybrid (*E. coli*)	*lacI*	IPTG	261, 263
P_{syn}, synthetic (*E. coli*)	*lacI*, *lacI*^{q}	IPTG	186
Starvation promoters (*E. coli*)			366
p_L (λ)	λ*cI*ts857	Thermal	43, 80, 129, 130, 240, 454
p_L-9G-50, mutant (λ)		Reduced temperature (<20°C)	187, 433
cspA (*E. coli*)		Reduced temperature (<20°C)	187, 206, 433, 551
p_R, p_L, tandem (λ)	λ*cI*ts857	Thermal	150, 493
T7 (T7)	λ*cI*ts857	Thermal	537, 548
T7-*lac* operator (T7)	*lacI*^{q}	IPTG	141, 190, 239
λp_L, p_{T7}, tandem (λ, T7)	λ*cI*ts857, *lacI*^{q}	Thermal, IPTG	375
T3-*lac* operator (T3)	*lacI*^{q}	IPTG	190, 605
T5-*lac* operator (T5)	*lacI*^{q}, *lacI*	IPTG	71, 390
T4 gene 32 (T4)		T4 infection	143, 210
nprM-lac operator (*Bacillus* spp.)	*lacI*^{q}	IPTG	605
VHb (*Vitreoscilla* spp.)		Oxygen, cAMP-CAP[e]	304, 305
Protein A (*Staphylococcus aureus*)			1,256, 349

[a] *lacI* gene with single mutation, Gly-187 → Ser (72).
[b] *lacI* gene with three mutations, Ala-241 → Thr, Gly-265 → Asp, and Ser-300 → Asn (604).
[c] The constitutive *lpp* promoter (P_{lpp}) was converted into an inducible promoter by insertion of the *lac*UV5 promoter/operator region downstream of P_{lpp}. Thus, expression occurs only in the presence of a *lac* inducer (142).
[d] Wild-type *lacI* gene.
[e] cAMP-CAP, cyclic AMP-catabolite activator protein.

Lanzer and Bujard carried out extensive studies on the commonly used *lac*-based promoter-operator systems and demonstrated up to 70-fold differences in the level of repression when the operator was placed in different positions within the promoter sequence (328). Thus, when the 17-bp operator was placed between the −10 and −35 hexameric regions, a 50- to 70-fold-greater repression was caused than when the operator was placed either upstream of the −35 region or downstream of the −10 site (328).

A third important characteristic of a promoter is its inducibility in a simple and cost-effective manner. The most widely used promoters for large-scale protein production use thermal induction (λ p_L) or chemical inducers (*trp*) (Table 1). The isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible hybrid promoters *tac* (123) or *trc* (65) are powerful and widely used for basic research. However, the use of IPTG for the large-scale production of human therapeutic proteins is undesirable because of its toxicity (159) and cost. These drawbacks of IPTG have until now precluded the use of the *tac* or *trc* promoter from the production of human therapeutic proteins and rendered the large-scale expression of proteins for basic research prohibitively expensive. The availability of a mutant *lacI*(Ts) gene that encodes a thermosensitive *lac* repressor (72) now permits the thermal induction of these promoters (4, 9, 234). In addition, the new vectors exhibit tight regulation of the *trc* promoter at 30°C (9). Two different *lac* repressor mutants that are thermosensitive (586, 604) as well as IPTG inducible (586) have recently been described. Although the wild-type *lacI* gene can be thermally induced (602, 603), this system is not tightly regulated and cannot be used in *lacI*^{q} strains, since a temperature shift does not override the tight repression caused by the overproduction of the *lac* repressor (603). Thus, this system is limited to the production of some proteins that are not detrimental to the host cell.

Cold-responsive promoters, although much less extensively studied than many of the other promoters included here, have been shown to facilitate efficient gene expression at reduced temperatures. The activity of the phage λ p_L promoter was highest at 20°C and declined as the temperature was raised (187). This cold response of the p_L promoter is positively regulated by the *E. coli* integration host factor, a sequence-specific, multifunctional protein that binds and bends DNA (164, 165, 188). The promoter of the major cold shock gene *cspA* (206, 551) was similarly demonstrated to be active at reduced temperatures (187). Molecular dissection of the *cspA* and p_L promoters led to the identification of specific DNA regions involved in the enhancement of transcription at lower temperatures; this has allowed the development of p_L derivatives that are highly active at temperatures below 20°C (433). The rationale behind the use of cold-responsive promoters for

Downloaded from http://mmbr.asm.org/ on October 5, 2015 by guest

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

gene expression is based on the proposition that the rate of protein folding will be only slightly affected at about 15 to 20°C, whereas the rates of transcription and translation, being biochemical reactions, will be substantially decreased. This, in turn, will provide sufficient time for protein refolding, yielding active proteins and avoiding the formation of inactive protein aggregates, i.e., inclusion bodies, without reducing the final yield of the target protein (433). It would be interesting to compare the transcriptional activities of other promoters derived from cold shock genes (288, 402).

Other promoters that have been characterized recently (Table 1) possess attractive features and should provide additional options for high-level gene expression systems. For example, the pH promoter (102, 561) is very strong: recombinant proteins are produced at levels of up to 40 to 50% of the total cellular protein (480). This expression level, however, will probably vary for different genes, because protein synthesis depends on translational efficiency as well as promoter strength.

E. coli promoters are usually considered in terms of a core region composed of the −10 and −35 hexameric sequences including a 15- to 19-bp spacer between the two hexamers (344). However, it has been proposed that elements outside the core region stimulate promoter activity (134). Many studies have demonstrated that sequences upstream of the core promoter increase the rate of transcription initiation in vivo (172, 213, 264, 290, 618). Gourse and colleagues have shown that a DNA sequence, the UP element, located upstream of the −35 region of the *E. coli* rRNA promoter *rrnB* P1, stimulates transcription by a factor of 30 in vitro and in vivo (290, 453, 468). The UP element functions as an independent promoter module because when it is fused to other promoters such as *lacUV5*, it stimulates transcription (453, 468). Upstream activation in *E. coli* and other organisms has been reviewed in detail (110). The ability of the UP element to act as a transcriptional enhancer when fused to heterologous promoters may be of general utility in high-level expression systems.

Although the extraordinary strength of the rRNA promoters P1 and P2 is well documented (173, 414), these promoters have not been exploited for the high-level production of proteins in *E. coli*, mainly because their regulation is more difficult. The in vivo synthesis of rRNA is subject to growth rate control (213), and P1 and P2 are active during periods of rapid cell growth and are downregulated when cells are in the stationary phase of growth. Therefore, the rRNA promoters would be continuously active or "leaky" during the preinduction phase. In vivo P2 is the weaker, less inducible promoter in rapidly growing cells. However, when uncoupled from P1, the P2 promoter shows increased activity (up to 70% of that of P1) and becomes sensitive to the stringent response, indicating that in its native tandem context, P2 is partially occluded (173, 289). Brosius and Holy (66) inserted the *lac* operator sequence downstream of the *rrnB* rRNA P2 promoter and achieved repression of P2 in strains harboring the *lacI*^{q} gene. Transcriptional activity was measured by the production of chloramphenicol acetyltransferase and by the expression of the 4.5S RNA. However, the P2 construction was only half as active as the *tac* promoter, and furthermore, when the *rrnB* P1 promoter was placed upstream of the P2 promoter, transcriptional repression was incomplete (66).

It is tempting to speculate that rRNA promoters could be tightly regulated by using the concept of inverted promoters (see the section on tightly regulated expression systems, below). Thus, a rRNA promoter could be cloned upstream of the gene of interest but in the opposite transcriptional direction. The use of λ integration sites and a regulated λ integrase would facilitate the inversion of the promoter for induction, and the presence of strong transcription terminators upstream of the highly active promoter would prevent destabilization of the vector during the preinduction phase.

Transcriptional Terminators

In prokaryotes, transcription termination is effected by two different types of mechanisms: Rho-dependent transcription termination depends on the hexameric protein rho, which causes the release of the nascent RNA transcript from the template. In contrast, rho-independent termination depends on signals encoded in the template, specifically, a region of dyad symmetry that encodes a hairpin or stem-loop structure in the nascent RNA and a second region that is rich in dA and dT and is located 4 to 9 bp distal to the dyadic sequence (83, 122, 439, 455, 456, 465, 594, 609). Although often overlooked in the construction of expression plasmids, efficient transcription terminators are indispensable elements of expression vectors, because they serve several important functions. Transcription through a promoter may inhibit its function, a phenomenon known as promoter occlusion (5). This interference can be prevented by the proper placement of a transcription terminator downstream of the coding sequence to prevent continued transcription through another promoter. Similarly, a transcription terminator placed upstream of the promoter that drives expression of the gene of interest minimizes background transcription (413). It is also known that transcription from strong promoters can destabilize plasmids as a result of overproduction of the ROP protein involved in the control of plasmid copy number as a result of transcriptional readthrough into the replication region (539). In addition, transcription terminators enhance mRNA stability (237, 404, 597) and can substantially increase the level of protein production (237, 572). Particularly effective are the two tandem transcription terminators T1 and T2, derived from the *rrnB* rRNA operon of *E. coli* (67), but many other sequences are also quite effective.

Transcriptional Antiterminators

In bacteria, many operons involved in amino acid biosynthesis contain transcriptional attenuators at the 5′ end of the first structural gene. The attenuators are regulated by the amino acid products of the particular operon. Thus, the availability of the cognate charged tRNA leads to the formation of a secondary structure in the nascent transcript followed by ribosome stalling. In the absence of the cognate charged tRNA, an antiterminator structure which prevents formation of the RNA hairpin in the terminator and prevents transcriptional termination is formed (325). The antiterminator element that enables RNA polymerase to override a rho-dependent terminator in the ribosomal RNA operons has been identified and is referred to as *boxA* (41, 341). Transcriptional antitermination is a remarkably complex process that involves many known and as yet unidentified host factors. This topic has been covered in great detail in two excellent recent reviews (110, 456). Here, we will briefly consider the use of antitermination elements that are useful in the expression of heterologous genes in *E. coli.*

One of the more powerful and widely used expression systems in *E. coli* makes use of the phage T7 late promoter (537, 548). The activity of this system depends on a transcription unit that supplies the T7 RNA polymerase, whose tight repression is essential to avoid leakiness of the T7 promoter. Several approaches have been used to regulate the expression of the T7 polymerase, and each has its own unique disadvantages (374). Mertens et al. (374) addressed this problem by constructing a reversibly attenuated T7 RNA polymerase expres-

Downloaded from http://mmbr.asm.org/ on October 5, 2015 by guest

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

sion cassette based on λ p_L regulation. Thus, the basal expression level of the T7 polymerase was attenuated by inserting three tandemly arranged transcription terminators between the promoter and the gene encoding the T7 polymerase. For induction, the phage λ-derived nut_L-dependent antitermination function was also incorporated to override the transcription block. Alternatively, an IPTG-inducible promoter was similarly used, allowing conditional reversion of attenuation upon induction (374).

The transcriptional antitermination region from the *E. coli rrnB* rRNA operon has been used in the expression vector pSE420, which utilizes the *trc* promoter (64). The rationale in this case was to facilitate transcription through areas of severe secondary structure, thus reducing the possibility of premature transcription termination by the host RNA polymerase. In this case, however, the presence of the *rrnB* antiterminator is apparently ineffective (64a).

Tightly Regulated Expression Systems

The advantages of tightly regulated promoters (see the section on promoters, above) have led to the design of many ingenious and highly repressible expression systems that are particularly useful for the expression of genes whose products are detrimental to host growth. The various approaches include the use of a "plating" method (544), the increase of the repressor-to-operator ratio (9, 391), induction by infection with mutant phage (68, 137), attenuation of promoter strength on high-copy-number vectors (587), the use of transcription terminators (374, 375, 413) in combination with antiterminators (374), the use of an inducible promoter within a copy-number-controllable plasmid (558), "cross-regulation" systems (97, 98), cotransformation of plasmids utilizing the SP6 RNA polymerase (473), and the use of antisense RNA complementary to the mRNA of the cloned gene (423). Finally, one elegant approach involves the principle of invertible promoters: the promoter, flanked by two λ integration sites, faces in the direction opposite that of the gene to be expressed and is inverted only by inducing site-specific genetic recombination mediated by the λ integrase (16, 21, 235, 441, 599).

The above systems have advantages as well as disadvantages, depending on their intended use. Thus, methods that rely on solid media cannot easily be used for large-scale expression. High-level repressor systems often cause a substantial decrease in protein yield (9, 531), thus necessitating optimization of the repressor-to-operator ratio (234). Induction mediated by λ phage adds further complexity to the system. The use of inverted promoter circuits involves complex vector constructions. Although most of the above systems have not yet been used for the high-level production of proteins on a large scale, they nevertheless provide important tools for the armamentarium of gene expression.

TRANSLATIONAL REGULATION

mRNA Translational Initiation

The extensive knowledge of the transcriptional process has allowed the use of prokaryotic promoters in cassette fashion, unaffected by the surrounding nucleotide context (232, 236, 317, 344). However, the determinants of protein synthesis initiation have been more difficult to decipher; this is not surprising, considering the complexity of this process (224, 579). It is now clear that the wide range of efficiencies in the translation of different mRNAs is predominantly due to the unique structural features at the 5′ end of each mRNA species. Thus, in contrast to the portable promoters, no universal sequence for the efficient initiation of translation has been devised. However, progress in this aspect of gene expression in *E. coli* has been strong, and general "guidelines" have emerged (131, 133, 196, 198, 218, 368, 369, 458, 579, 590).

The translational initiation region of most sequenced *E. coli* genes (91%) contains the initiation codon AUG. GUG is used by about 8% of the genes, and UUG is rarely used as a start site (1%) (218, 224, 535). In one case, AUU is used as the start codon for *infC* (75). This codon is required for the autogenous regulation of *infC*. The translational efficiency of the initiation codons in *E. coli* has been examined. AUG is the preferred codon by two- to threefold, and GUG is only slightly better than UUG (458, 573).

Shine and Dalgarno (514, 515) identified a sequence in the RBS of bacteriophage mRNAs and proposed that this region, subsequently called the Shine-Dalgarno (SD) site, interacts with the complementary 3′ end of 16S rRNA during translation initiation. This was confirmed by Steitz and Jakes (532). The spacing between the SD site and the initiating AUG codon can vary from 5 to 13 nucleotides, and it influences the efficiency of translational initiation (196). Extensive studies have been carried out to determine the optimal nucleotide sequence of the SD region, as well as the most effective spacing between the SD site and the start codon (28, 93, 131, 593). Ringquist et al. (458) examined the translational roles of the RBS and reached the following conclusions. (i) The SD sequence UAA GGAGG enables three- to sixfold-higher protein production than AAGGA for every spacing. (ii) For each SD sequence, there is an optimal although relatively broad spacing of 5 to 7 nucleotides for AAGGA and 4 to 8 nucleotides for UAAGG AGG. (iii) For each SD sequence, there is a minimum spacing required for translation; for AAGGA, this minimum spacing is 5 nucleotides, and for UAAGGAGG, it is 3 to 4 nucleotides. These spacings suggest that there is a precise physical relationship between the 3′ end of 16S rRNA and the anticodon of the fMet-$tRNA_f$ bound to the ribosomal P site (458).

The secondary structure at the translation initiation region of mRNA plays a crucial role in the efficiency of gene expression (132, 229, 233, 277, 295). It is believed that the occlusion of the SD region and/or the AUG codon by a stem-loop structure prevents accessibility to the 30S ribosomal subunits and inhibits translation (184, 451, 556). Several different strategies have been devised to minimize mRNA secondary structure. The enrichment of the RBS with adenine and thymidine residues enhanced the expression of certain genes (94, 412, 429). Similarly, the mutation of specific nucleotides upstream or downstream of the SD region suppressed the formation of mRNA secondary structure and enhanced translational efficiency (107, 223, 266, 336, 530, 583). Another approach takes advantage of the naturally occurring phenomenon of translational coupling in bacteria (506). The mechanism of translational coupling has been invoked to account for the coordinate expression of different proteins from polycistronic mRNAs. Thus, it was shown that the moderately strong *gal* promoter could direct the synthesis of galactokinase at very high levels when *galK* was translationally coupled to an upstream gene, suggesting that even a weak RBS may be highly efficient if it is accessible to ribosomes (506). Schümperli et al. (506) suggested that this regulatory mechanism might have important applications in biotechnology for the overproduction of proteins. Indeed, translational coupling has been widely used for the high-level expression of diverse genes (46, 359, 430, 438, 503, 504, 505, 552).

In addition to the binding of the SD region to the 16S rRNA, other interactions between mRNA and the ribosome are in-

Downloaded from http://mmbr.asm.org/ on October 5, 2015 by guest

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.