Volume 17 Number 8 1989 Nucleic Acids Research

An additional ribosome-binding site on mRNA of highly expressed genes and a bifunctional site on the colicin fragment of 16S rRNA from *Escherichia coli*: important determinants of the efficiency of translation-initiation

T.A.Thanaraj and M.W.Pandit*

Centre for Cellular and Molecular Biology, Hyderabad 500 007, India

Received February 1, 1989; Revised and Accepted March 23, 1989

ABSTRACT

For various genes of *E.coli*, three regions (−55 to −1; −35 to −1; −21 to −1) 5′ to AUG codon on mRNA were searched for sites of interaction with colicin fragment of 16S rRNA. The detailed sequence comparison points out that apart from Shine−Dalgarno base pairing, an additional ribosome-binding site, a subsequence of 5′-UGAUCC-3′ invariably exists in mRNA for highly expressed genes. Poorly expressed genes appear to be controlled by only Shine−Dalgarno base pairing. The analysis indicates that in the initiator region, the −55 to −1 region contains the signal which decides the efficiency of the translation-initiation. The site on 16S rRNA, 5′-GGAUCA-3′ at position 1529, that can base pair to the above site, has a recognition site on 23S rRNA at position 2390. In the light of the conserved nature and accessibility of these sites, it is proposed that the site on 16S rRNA plays a bifunctional role—initially it binds to mRNA from highly expressed genes to form a stable 30S initiation complex, and upon association with 50S subunit it exchanges base pairing with 23S rRNA, thus leaving the site on mRNA free.

INTRODUCTION

The mechanisms of information transfer from gene to protein and the role of controls exerted by these mechanisms on the level of expression have been the subject of many experimental and theoretical studies. In prokaryotes, the major controls are exerted at the levels of transcription, half life of mRNA, and translation. The controls for translational efficiency are exerted through processes such as initiation (1−4), elongation (5−7) and termination (8). The relative rate of biosynthesis of various proteins is known to vary with one or more controlling factors mentioned above and the cellular physiology of the organism. The steady-state level of various proteins again depends on their rates of turnover.

It has been recognised (9) that the initiation of translation is one of the major rate-limiting steps in protein synthesis under normal situations. The translation-initiation process involves binding of mRNA and initiator tRNA to the 30S ribosomal subunit (10) facilitated by the initiation factors (11); the 30S initiation complex in turn binds to 50S subunit leading to the 70S initiation complex as the end product (12). The rate of formation of the 30S initiation complex largely depends on the ribosome-binding strength (due to Shine−Dalgarno (SD) base pairing (13) and AUG codon binding) which is modulated by the spacing between the SD sequence and AUG (12,14), and by secondary structure in the initiation region (15−16).

Even though the ribosome-protected segment of the mRNA was found to contain 20 to 25 nucleotides 5′ to AUG codon, it has been noticed by several workers that important determinants of ribosome-binding strength lie 5′ to the SD region (17−18). Isolated ribosome-binding sites including SD region and AUG codon do not usually rebind to ribosomes (14). *In vitro* experiments by Borisova *et al.* (19) on systematic rebinding of ribosomes with the MS2 transcript for replicase indicated that the −53 to +6 region retained most of the activity (> =60%); −35/−33 to +6 region retained 42% of the activity, and −21 to +6 region showed no activity. An *in vivo* study by Kastelein *et al.* (20), has shown that successive gradual erasure of the 5′ terminal sequence of MS2 DNA for the coat protein, in the region beyond 30 bases upstream of AUG codon profoundly reduces translation. Thus, it appears that there are extra sequences upstream of the ribosome-protected segment that make, with the incoming 30S ribosome, transient contacts which are not maintained in the mature 70S initiation complex. Similar idea that essential features of the ribosome-binding lie outside the ribosome-protected segment is supported by other workers also (21). In this paper, we report the results of an analysis of the sequence upstream of AUG in mRNA, which suggests the existence of a defined distant sequence on mRNA that contributes significantly to the ribosome-binding strength and contains the information responsible for the efficiency of the translation-initiation in *E.coli*. We also propose a mechanism whereby the additional binding sites, whenever they exist, are released upon association of the 50S subunit.

METHOD

In our study, we have considered mainly RNA−RNA interactions, as it has been observed that mRNA−rRNA interactions are enough to maintain mRNA−rRNA hybrid (14) and that proteins play no essential role in the modulation of the efficiency of translation initiation (19). Gouy and Gautier (5) have reviewed and listed the cellular contents of various *E.coli* gene products for their study on codon usage and gene expressivity. This data is based on the steady state levels of proteins arising out of various controls in protein synthesis, and does not correlate these yields to specific steps involved in protein synthesis. In the absence of other data on the relative yields of translation, we used the above data for our study. For each gene of known protein yield, we have considered three regions: (i) −21 to −1 (small region) which is the ribosome-protected segment, (ii) −35 to −1 (medium region), and (iii) −55 to −1 (large region), for reasons mentioned in the introduction. As it is known that translational efficiency is inversely proportional to the degree of secondary structure at the 5′ noncoding region (22), we thought it necessary to take into account the secondary structure of mRNA regions. Using our program package COMPLEMENT (23) reported earlier, the intracomplementary stretches within each type

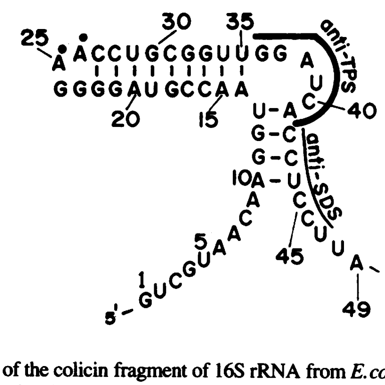

Figure 1. Secondary structure of the colicin fragment of 16S rRNA from *E.coli*, showing the anti-TP site (36−41) and the anti-SD site (42−47). Methylated bases are indicated by dots.

of mRNA region were located which can give the secondary structure of the region. The secondary structure of the colicin fragment of 16S rRNA (49 nucleotides at the 3′ end) is known (24), and is shown in Figure 1. With the use of COMPLEMENT program, all possible intermolecular complementary stretches between each mRNA region and the colicin fragment of 16S rRNA were located. The sites thus obtained were subjected to the following criteria in order to pick up the potential sites of interaction for ribosome binding.

(i) Each site should be comprised of a contiguous sequence of bases and should involve only Watson and Crick base pairing.

TABLE 1 - Shine-Dalgarno base pairing in -55 to -1 region

Gene	Site on mRNA		Site on 16S rRNA		$-\Delta G$ Kcal/mol	ΔM	ΔR
Highly Expressed Genes							
lpp	-13 to -10	--GAGG	42 to 45	CCUC--	6.9	9	4
tufB	-12 to -9	--GAGG	42 to 45	CCUC--	6.9	8	4
recA	-10 to -7	AGGA--	44 to 47	--UCCU	6.9	6	2
ompA	-12 to -9	--GAGG	42 to 45	CCUC-	6.9	8	4
tsf	-10 to -7	AGGA--	44 to 47	--UCCU	6.9	6	2
rplK	-11 to -8	AGGA--	44 to 47	--UCCU	6.9	7	2
rplC	-12 to -8	--GAGGU	41 to 45	ACCUC--	9.0	7	4
rpmH	-14 to -11	---AGGU	41 to 44	ACCU---	6.7	10	5
rpsU	-13 to -11	---AGG	42 to 44	CCU---	4.6	7	5
rpsM	-11 to -7	AGGAG-	43 to 47	-CUCCU	8.6	6	2
rplQ	-11 to -8	AGGA--	44 to 47	--UCCU	6.5	7	2
rpsJ	-14 to -11	-GGAG-	43 to 46	-CUCC-	6.9	10	3
rplD	-11 to -7	AGGAG-	43 to 47	-CUCCU	8.6	6	2
rplB	-14 to -10	-GGAGG	42 to 46	CCUCC-	9.8	9	3
rpsS	-11 to -8	--GAGG	42 to 45	CCUC--	6.9	7	4
rplV	-11 to -6	AGGAGG	42 to 47	CCUCCU	11.5	5	2
rpsC	-11 to -8	-GGAG-	43 to 46	-CUCC-	6.9	7	3
rpsO	-11 to -8	-GGAG-	43 to 46	-CUCC-	6.9	7	3
rpsQ	-13 to -11	AGG---	45 to 47	---CCU	4.6	10	2
				Average $-\Delta G$	7.3		
Moderately Expressed Genes							
lacY	-12 to -9	AGGA--	44 to 47	--UCCU	6.9	8	2
trpA	-13 to -10	--GAGG	42 to 45	CCUC--	6.9	9	4
trpB	-12 to -9	AGGA--	44 to 47	--UCCU	6.9	8	2
trpC	-11 to -8	--GAGG	42 to 45	CCUC--	6.9	7	4
trpD	-12 to -9	-GGAG-	43 to 46	-CUCC-	6.9	8	3
trpE	-10 to -8	-GGAG-	43 to 46	-CUCC-	5.2	7	3
rpoB	-10 to -7	--GAGG	42 to 45	CCUC--	6.9	6	4
				Average $-\Delta G$	6.6		
Poorly Expressed Genes							
trpR	-9 to -6	(CGAC)	1 to 4	(GUCG)	6.4	5	45
dnaG	-10 to -8	(AUC)	37 to 39	(GAU)	3.2	7	10
eltA	-13 to -11	AAG----	46 to 48	----CUU	2.6	10	1
galR	-12 to -10	UAA-----	47 to 49	-----UUA	1.8	9	0
lexA	-3 to -1	-GGA--	44 to 46	--UCC-	5.2	0	3
				Average $-\Delta G$	4.3		

The subsequences not found to be a part of either SD or anti-SD sequence are given in brackets. The dash (-) indicates the base other than the corresponding base in SD or anti-SD as the case may be.

TABLE 2 - Additional binding sites in three regions

Gene	Site on mRNA		Site on 16S rRNA		- ΔG Kcal/ mol
	position	sequence UGAUCC*	position 36 to 41	sequence GGAUCA*	
I. -21 to -1 Region					
Highly Expressed Genes					
ompA	-19 to -16	UGAU--	38 to 41	--AUCA	5.0
rplD	-22 to -19	GUGA---	39 to 42	---UCAC	6.2
Moderately Expressed Genes					
lacY	-7 to -4	---UCCA	35 to 38	UGGA---	5.1
trpD	-8 to -6	(GAC)	1 to 3	(GUC)	3.8
Poorly Expressed Genes					
eltA	-20 to -17	UGAU--	38 to 41	--AUCA	5.0
II. -35 to -1 Region					
Highly Expressed Genes					
tufB	-24 to -21	--AUCC	36 to 39	GGAU--	6.1
recA	-32 to -29	--AUCC	36 to 39	GGAU--	5.1
ompA	-19 to -16	UGAU--	38 to 41	--AUCA	5.0
tsf	-23 to -20	-GAUC-	37 to 40	-GAUC-	5.5
rplK	-27 to -24	(GUUA)	5 to 8	(UAAC)	3.9
rplD	-22 to -19	UGA---	39 to 41	---UCA	6.2
rpsC	-26 to -23	-GAUC-	37 to 40	-GAUC-	5.5
rpsO	-31 to -28	----CCUU	9 to 12	AAGG----	5.5
Moderately Expressed Genes					
lacY	-7 to -4	--AUCC	36 to 39	GGAU--	6.1
trpA	-15 to -12	(CGA)	2 to 5	(UCG)	6.4
Poorly Expressed Genes					
eltA	-26 to -22	(CUUGU)	7 to 11	(ACAAG)	6.5
galR	-22 to -18	(UAAGG)	45 to 49	(CCUUA)	6.6
III. -55 to -1 Region					
Highly Expressed Genes					
lpp**	-27 to -25	-GAU--	38 to 40	--AUC-	3.2
tufB	-39 to -36	UGAU--	38 to 41	--AUCA	6.1
recA	-32 to -29	--AUCC	36 to 39	GGAU--	6.1
ompA	-19 to -16	UGAU--	38 to 41	--AUCA	5.0
tsf	-23 to -20	-GAUC-	37 to 40	-GAUC-	5.5
rplK**	-49 to -47	--AUC-	37 to 39	-GAU--	5.2
rplC	-48 to -44	-GAUC-	37 to 40	-GAUC-	5.5
rpmH	-46 to -43	--AUCC	36 to 39	GGAU--	6.1
rpsU	-41 to -38	UGAU--	38 to 41	--AUCA	5.0
rpsM	-58 to -55	--AUCC	36 to 39	GGAU--	5.1
rplQ	-38 to -35	-GAUC-	37 to 40	-GAUC-	5.6
rpsJ	-38 to -36	--AUC-	37 to 39	-GAU--	3.2
rplD	-39 to -35	UGAUC-	37 to 41	-GAUCA	7.3
rplB	-46 to -44	--AUC-	37 to 39	-GAU--	3.2
rpsS	-50 to -47	UGAUC-	38 to 41	-GAUCA	5.0
rplV	-39 to -36	UGAUC-	38 to 41	-GAUCA	5.0
rpsC	-26 to -23	-GAUC-	37 to 40	-GAUC-	5.5
rpsO	-40 to -37	-GAUC-	37 to 40	-GAUC-	5.5
rpsQ	-47 to -45	-GAU--	38 to 40	--AUC-	3.2
			Average	- ΔG	5.1
Moderately Expressed Genes					
None					
Poorly Expressed Genes					
None					

*Hexamer sequences arising out of additional binding site;subsequences which are not part of these sequences are given in brackets;the dash(-) indicates the base other than the corresponding base in the hexamer sequences (*)

**These genes have one more additional site as given below

lpp	-45 to -42	(CUUGU)	7 to 11	(ACAAG)	65
rplK	-15 to -12	(CUUG)	8 to 11	(CAAG)	44

(ii) Only tetranucleotides and higher-order stretches were considered first; however, in the absence of these, trinucleotide stretches were considered.

(iii) If a stretch from anti-SD region of colicin fragment has another complementary site on mRNA in addition to SD site, the SD base pairing was preferred.

(iv) Intermolecular complementary stretches whose corresponding sites on colicin fragment fall either in the central hairpin stem or in the central hairpin loop were ignored for the reasons that the above central stem is highly stable and the methylated adenosines in the above central hairpin loop are known to play a role in binding fMet−tRNA (25).

(v) Intermolecular complementary stretches whose corresponding sites on mRNA region are involved in intramolecular base pairing, were ignored if (1) such a base pairing was of equal or higher order, or (2) > = 50% of the bases were involved in the equal or higher-order base pairing, or (3) < 50% of the bases were involved in the higher-order base pairing. If < 50% of the bases in the site are involved in base pairing of the same order, the contribution from free energy is taken into account in order to decide whether the preferred base pairing will be intra- or intermolecular.

Such selected intermolecular complementary sites were classified as Shine−Dalgarno binding site and additional binding site. Additional binding sites were those that did not fall in the SD region. The free energy of such sites was calculated (26) and the spacings for the SD base pairing in mRNA as well as in the colicin fragment were identified. It has been found that a space of 5−10 nucleotides between AUG codon and SD sequence on mRNA and a space of 2−5 nucleotides between the anti-SD sequence and the 3′ terminal end of 16S rRNA are in the acceptable range (27) and that extending the spacing outside the acceptable range has a deleterious effect (12). A larger number of continuous base pairing in the SD region led to higher efficiency of translation-initiation (13). In view of the above mentioned observations, we distinguish well-defined SD base pairing using the criteria that a well-defined SD base pairing should (i) either involve 4 or more continuous bases, or have $-\Delta$ G> =4.0 Kcal/mol, (ii) have the spacing (ΔM) on mRNA within 5−10 and the spacing (ΔR) on 16S rRNA within 2−5, and (iii) be free from intramolecular secondary structure. If even one of these criteria was violated for an SD base pairing, such an SD base pairing is referred to as ill-defined.

RESULTS AND DISCUSSION

The method discussed above was applied to various genes of known expressivity (5). The results are summarised in Tables 1 and 2. Based on the steady-state concentration of genome products, genes analysed in this study were classified into three groups, namely, (i) highly expressed genes (HE-genes), for which gene products exceed 9000 mol/genome, (ii) moderately expressed genes (ME-genes), for which gene products range between 1400−3500 mol/genome, and (iii) poorly expressed genes (PE-genes), for which gene products are < =100 mol/genome.

The SD base pairing in the case of −55 to −1 region is given in Table 1. The ill-defined SD base pairings are indicated by underlining the values for ΔG, ΔM and ΔR in Table 1. As can be seen from the Table, all of the HE- and ME-genes have well-defined SD base pairing where as PE- genes have ill-defined SD base pairing. This pattern was found to be more or less the same even when shorter regions (−35 to −1, and −21 to −1) were analysed, except that in the case of −21 to −1 region, the gene *rpsU* of HE-class had ill-defined base pairing(ΔM=2) and in the case of −35 to -1 region, the gene *rpoB* of ME-class had ill-defined base pairing (ΔM=20 and ΔR=38). These ill-defined base

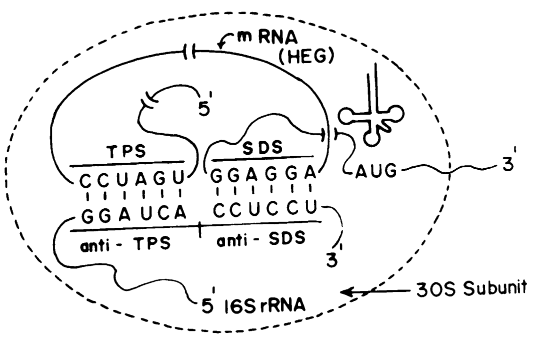

Figure 2. Schematic representation of the 30S initiation complex formation, showing the TP site base pairing for HE-mRNAs.

pairings turned into well-defined ones, when the bigger regions were considered, indicating that the sites which were involved in intramolecular secondary structure have now become free. The averages of free energy (Kcal/mol) due to SD base pairing for HE-, ME- and PE-genes were found to be −7.3, −6.6 and −4.3 respectively. In fact, the ΔG (−4.3) for PE-genes will be reduced further, if properly corrected for its ill-defined SD base pairing. The trend of these values remained the same irrespective of the length of the region analysed.

The search for the additional sites showed that unlike SD site, these did not occur on each and every gene that we examined. In Table 2, the additional sites of interaction that occurs in the three types of regions, are given. Only those genes for which such sites exist are included in the table. As can be seen from the table, in the case of small regions we could pick up a few genes from each class in which an additional site was found; the existence of such a site appeared, therefore, to be independent of the level of the gene expression. In the case of medium regions, the number of HE-genes showing the existence of the additional site was appreciably higher (8 out of 19 HE-genes); this suggested that the existence of additional site of interaction may be coupled with the high expression of gene. Finally, when the large region was considered, it was found that the additional site existed in all the HE-genes. It is interesting to note that the additional sites that existed in the medium as well as the small regions for ME- and PE-genes have not been picked up in the case of the large regions obviously because they possess intramolecular complementary stretches in the extended region (−36 to −55). The HE-genes which did not have additional sites in the medium region have now acquired it, presumably as a consequence of the availability of the extended region; only one exception to this was the gene *lpp*, the site on which was missed earlier due to its involvement in intramolecular base pairing and is now picked up because of the rearrangement of the secondary structure of the extended region. The average −ΔG (5.1) Kcal/mol arising out of the additional site can lead up to a 1000-fold increase in the association strength (12).

Due to the contribution from the additional site in the case of HE-genes, the available energy is expected to increase to the level of −12.4. Hence, the free energy which is a measure of the ribosome-binding strength, could be directly correlated with the level of expression of the gene. The additional ribosome-binding site, therefore, can promote/enhance the process of translation-initiation; we call this site as TP (Translation-initiation Promoting) site. A schematic diagram of the 30S initiation complex including the TP site is given in Figure 2.

TABLE 3 - Results of the search for SD and TP sites on other mRNA genes of E.coli

Gene	SDS	TPS	Gene	SDS	TPS
A.Genes of known high expression or high translation efficiency					
rplL	AGGA--(-13 to -10)	UGAU--(-21 to -18)	tufA	AGGA--(-12 to -9)	---UCC(-37 to -35)
lamB	AGGAG-(-10 to -6)	UGAU--(-35 to -32)	bla	AGGA--(-9 to -6)	--AUCC(-59 to -56)
melA	AGGAG-(-11 to -7)	UGAU--(-34 to -31)	malG	-GGAG-(-13 to -10)	UGAU--(-21 to -18)
papH	-GGAG-(-10 to -7)	--AUCC(-55 to -52)	papC	AGGAG-(-11 to -7)	--AUCC(-59 to -56)
papB	--GAGG(-11 to -8)	-GAUC-(-54 to -51)	rpsD	--GAGG(-9 to -6)	UGAUC-(-51 to -47)
rpmD	-GGA--(-11 to -9)	--AUCC(-31 to -28)	papD	---AGG(-7 to -5)	UGAUC-(-79 to -75)
papE	-GGAG-(-10 to -7)	-GAUC-(-84 to -78)	rplF	-GGA--(-10 to -8)	UGAUC-(-70 to -66)
papF	--GAGG(-13 to -10)	UGAU--(-81 to -78)	rpsH	-GGAGG(-14 to -10)	--AUCC(-65 to -62)
malB	not well-defined	--AUCC(-38 to -35)	rpsA	not well-defined	--AUCC(-48 to -45)
rpsK	not well-defined	-GAUC-(-29 to -26)			
B.Genes of known moderate expression or moderate translation efficiency					
glys	--GAGG(-9 to -6)	------ ----------	atpE	-GGAG-(-10 to -7)	----- ----------
atpF	--GAGG(-13 to -10)	------ ----------	atpH	AGGAGG(-15 to -10)	----- ----------
C.Genes of known poor expression or poor translation efficiency					
eltB	-GGA--(-9 to -7)	------ ----------			
rieM	---AGG(-12 to -10)	------ ----------			
rieN	-GGA--(-14 to -12)	------ ----------			
lysR	--GAG-(-12 to -10)	------ ----------			

In all the four genes of category C, the SD sites are involved in intramolecular secondary structure.

TABLE 4 - Results of the search for TP sites on the mRNA genes from chloroplasts, other prokaryotes, and E.coli plasmids

Gene	TP Site	Gene	TP Site
Highly Expressed Genes			
Tobacco chloroplast		Other prokaryotes	
rpl33	--AUCC(-15 to -12)	M. capricolum	
rps7	UGAU--(-25 to -22)	rpsH	UGAU--(-27 to -24)
rps11	--AUCC(-40 to -37)	rplF	UGAU--(-19 to -16)
rps15	--AUCC(-38 to -35)	B. stearothermo-	
rps8	UGAU--(-42 to -39)	philus	
rps3	UGAU--(-24 to -21)	IF2	-GAUC-(-21 to -18)
rps22	UGAU--(-26 to -23)	E. aerogens	
rps7	UGAU--(-52 to -49)	ompA	UGAU--(-18 to -15)
Plasmids		B. cereus	
F-traa	--AUCC(-20 to -17)	bla	UGAU--(-58 to -55)
F-protein H	UGAU--(-39 to -36)	K. pneumoniae	
F-protein D	UGAU--(-55 to -52)	lacZ (beta	--AUCC(-44 to -41)
F-protein B	-GAUC-(-35 to -32)	galactosidase	
col ib-p9	--AUCC(-59 to -57)	**Moderately and Poorly Expressed Genes**	
R386-propilin	--AUCC(-20 to -17)	None of the following genes have TP	
cole3-ca38	-GAUCC(-42 to -38)	sites: trpE (B. amyloliquefaciens);	
colE1-imm	-GAUCC(-41 to -37)	trpC, trpD, rpoD, (B. subtilis);	
colE3-B	-GAUCC(-41 to -37)	lacY, lacI, (K.pneumoniae); blm	
		(B.cereus)	

Thus it appears that it is essential to examine the large region (−55 to −1) in order to decipher the complete information that can manifest in the formation of SD and TP base pairings and can explain the differential expression of genes through the promotion of translation-initiation.

Translation-initiation Promoting Site

As can be seen from Table 2, the TP sites on mRNAs and the complementary (anti-TP) sites on 16S rRNA, arise from defined hexamer sequences. TP site has a 5′-UGAUCC-3′ sequence and anti-TP a 5′-GGAUCA-3′ sequence. Moreover, the anti-TP site on the colicin fragment is found at the position 36 to 41, and is located 5′ to the anti-SD site at the position 42−47 (Fig. 1). The anti-TP site on 16S rRNA occurs as a small loop (at position 36 to 40) between the central hairpin stem and the additional weaker secondary structure in the colicin fragment as given in Figure 1. Even though the TP site on mRNA arises from a defined sequence, its position is not fixed. In this respect, TP base pairing is similar to SD base pairing because it is known that SD base pairing arises from defined sequence at a fixed position on 16S rRNA but with slight variations in its position on mRNA. The existence of a uniform pattern of the TP site was revealed only when large regions were considered. This again emphasises the importance of the large region. Because of the importance of the proposed function of the anti-TP site, one would expect it to be conserved in various species. We, in fact, found that the anti-TP site is well conserved (discussed later). The single-stranded nature of the TP site, and the anti-TP site, the conserved nature of the anti-TP site, and the occurrence of the TP site, in all the HE-genes considered so far, strongly suggest that the TP site and the anti-TP site can interact, leading to a stable 30S initiation complex. This stable 30S complex can then lead to a higher rate of formation of the 70S initiation complex that can enhance the translational efficiency. It is interesting

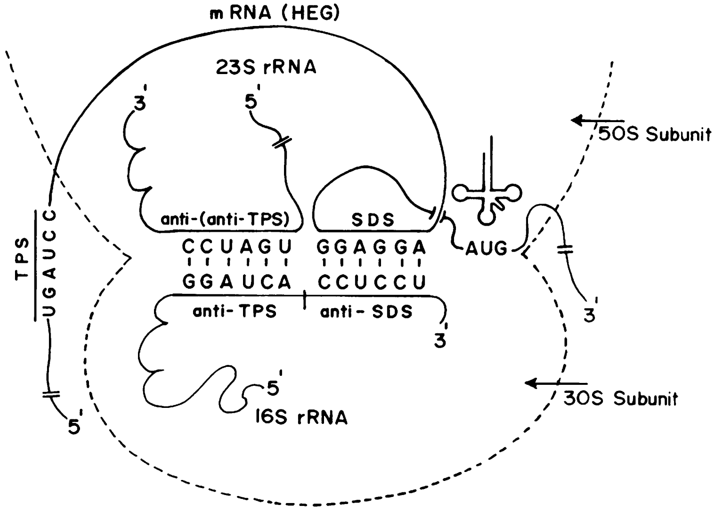

Figure 3. Schematic representation of the 70S initiation complex formation, showing interaction between the anti-TP site and the anti-(anti-TP) site, and interchange of base pairing for HE-mRNAs.

to note that in the Kastelein's experiments on MS2 *CP* gene (20) mentioned earlier, the region giving the maximum activity was −55 to +3, with the TP site existing at the position −49 to −45.

Having established the occurrence of the TP site on HE-genes and the necessity to take into account the large (−55 to −1) region to pick up the TP sites, we extended our analysis to other genes from *E.coli*. There are several genes in *E.coli* which are shown or concluded to be poorly or highly expressed. Some of them like *lamB*, *bla* (28), *pap* and *mal* (29) are known to have high translational efficiency even though the cellular content of these proteins is not known. There are several other genes of ribosomal proteins which we have not considered so far but that are also expected to be in the class of highly expressed genes. Atp mRNAs, which lead to high translational yield because of their high stability (30) can be classified as moderately expressed; they are not as highly expressed as the genes for ribosomal proteins. *rieM*, *rieN*, *lysR* and *eltb* are known to be poorly expressed (31). In the case of these genes, our analysis was restricted to large (−55 to −1) regions only and was limited to the search for the SD and the TP sites as defined earlier. These results are given in Table 3. It is interesting to note that in all the cases, TP site was found to exist, although in 5 out of the 20 cases it was found to be located outside the expected range. In 3 cases out of 20, SD site itself was not well-defined. Taking into account these deviations, the analysis supported, in general, our proposition that TP site exists only in those genes which are known for their high translational efficiency.

As the proposition about the existence of the TP site on HE-genes was supported by most of the tested *E.coli* genes, we examined the genes from chloroplasts, *E.coli* plasmids, and a few other prokaryotes. The results are given in Table 4. They not only provide additional support to our proposition but also show that occurrence of the TP site is a general feature in prokaryotes. This again suggests that SD base pairing is not the only mechanism for ribosome-binding but there exists the sequence (TP site) which can play the role of an additional ribosome-binding site. Our contention is also supported by similar observations in regard to the genes of lower eukaryotes such as yeast, which will be published separately.

Interchange of Base Pairing with 50S Subunit—Fate of the TP Site As can be seen from Tables 2 to 4, a major number of mRNA genes have the TP site outside the ribosome-protected segment. Considering the fact that only a region of 20–25 nucleotides 5′ to AUG codon is protected inside the ribosome, the proposed TP base pairing on 30S initiation complex has to be disrupted during the process of association of the 50S subunit. Such conformational change can be brought about by 16S rRNA interchanging its base pairing from mRNA to 23S rRNA upon the association of the 50S subunit, as shown in Figure 3. In the assembly of ribosomes, it has been observed that the association of 50S and 30S subunits takes place mainly through interactions between 16S rRNA and 23S rRNA (23,32). Keeping this in view, we searched the 23S rRNA for a site complementary to the anti-TP site. We found that there exists a unique complementary site 5′-UGAUCC-3′ on 23S rRNA at position 2390 which may serve as anti-(anti-TP) site. This site should exist in a single-stranded conformation and has to be accessible in order to form base pairs with the anti-TP site. Kethoxal modification studies by Herr & Noller (33, 34) have already shown that this site on 23S rRNA is accessible in the 50S subunit. It has also been shown that the corresponding site on 16S rRNA is kethoxal-reactive and that such chemical modification affects to some extent the formation of the 70S ribosome (35). Douthwaite *et al.* (36) have also shown that the concerned site on 16S rRNA is on the surface. It is worth noting that in the well accepted secondary structure model for 23S rRNA, proposed by Noller (37), except the 3′ base in the anti-(anti-TP) site, the other 5 bases are in a single-stranded form. Because of the functional role of these recognition sites between the 30S complex and

.TABLE 5 - Conservation of anti-TP site [5'-GGAUCA-3'] on 16S rRNA and anti-(anti-TP)site [5'-UGAUCC-3'] on 23S rRNA

Species	16S rRNA	23S rRNA
Eubacteria		
E.coli	GGUUGGAUCACCUC	AUAGUGAUCCGGUG
Mycoplasma capricolum	GGAUGGAUCACCUC	-
Anacystis nidulans	GGCUGGAUCACCUC	UUGAUGAUCCGACG
Proteus vulgaris	GGUUGGAUCACCUC	-
Pseudomonas aeruginose	-	AUAGUGAUCCGGUG
Chloroplast		
Maize	GGCUGGAUCACCUC	UUAGUGAUCCGACG
Tobacco	GGCUGGAUCACCUC	UUAGUGAUCCGACG
Euglena gracilis	GGCUGGAACAACUC	-
Chlamydomonas reinhardii	CCCUGGCUCACCUC	-
Archaebacteria		
Halobacterium volcanii	GGCUGGAUCACCUC	CUAGCGAACCAAUU
Halococcus morrhua	GGCUGGAUCACCUC	-
Methanococcus vannielii	GGCUGGAUCACCUC	CUAACGAACCCCUG
Sulfolbus solfatarious	GGCUGGAUCACCUC	-
Bacillus stearothermophilus	-	UUAGUGAUCCGGUG

Dash (-) indicates the non-availability of the sequence of the corresponding rRNA.

the 50S subunit, one would expect these sites to be conserved in various species. In order to study the conservation of these sites, we chose various species from evolutionarily distant organisms, such as, eubacteria, chloroplast, and archaebacteria for both 16S rRNA and 23S rRNAs. This data is given in Table 5. It can be seen that in most of the species analysed, the anti-TP site as well as the anti-(anti-TP) site is conserved. All these studies strongly suggest that the anti-TP and the anti-(anti-TP) sites may take part in the assembly of the 70S initiation complex.

The interchange of base pairing of similar type has been proposed by Van Duin (38) between 3′ ends of 16S and 23S rRNAs; but, on the basis of non-accessiblity of the concerned sites, Noller & others (33) did not find any evidence to support the model. However, the anti-TP and the anti-(anti-TP) sites proposed by us have been found to be conserved and experimentally free in the respective subunits. The anti-(anti-TP) site has been found to be equally accessible for kethoxal modification in 70S ribosome (33); this could be explained by the fact that these studies were performed on 70S ribosomes and not on 70S initiation complexes.

The mechanism involved in the proposed interchange of base pairing and the factors involved in such a process, are not clear at present. Initiation factor (IF3) is known to be responsible for the maintenance of the 30S and the 50S ribosomal subunits in a dissociated form; this factor is also involved in the binding of mRNA to the 50S subunit. These observations are suggestive of a possible mechanism which can involve IF3, and through

Table 6. mRNA genes from E.coli. predicted to have a high rate of translation-initiation

Gene	SDS (Location) AGGAGG	TPS (Location) UGAUCC
Ada - regulatory protein of adaptive response to alkylating agent	-GGAG-(-10 to -7)	--AUCC(-22 to -19)
pacI - penicillinase	-GGAG-(-13 to -10)	-GAUC-(-37 to -34)
galK - galaktokinase	-GGAG-(-12 to -9)	--AUCC(-30 to -27)
ilvE - branched chain amino-acid aminotransferase	AGGA--(-12 to -9)	--AUCC(-31 to -28)
nrdB - B2 subunit/ribonucleoside diphosphate reductase	AGGA--(-14 to -11)	--AUCC(-35 to -32)
phoT (pstA) - phosphate transport	--GAGG(-14 to -11)	UGAU--(-37 to -34)
ponB - penicillin-binding protein 1B (small component)	AGGA--(-14 to -11)	UGAU--(-25 to -22)
proA - gamma-glutamyl phosphate reductase (GPR)	AGGAG-(-12 to -8)	UGAU--(-25 to -22)
pyrD - dihydrooratate dehydrogenase	AGGA--(-10 to -7)	--AUCC(-14 to -11)
pyrE - orotate phosphoribosyl-transferase	AGGAG-(-12 to -8)	--AUCC(-52 to -49)
tagI - 3-methyladenine DNA glycosylase	--GAGG(-10 to -7)	-GAUC-(-25 to -22)
tnaa - tryptophanase	AGGA--(-11 to -8)	-GAUC-(-35 to -32)
gidA	--GAGG(-12 to -9)	--AUCC(-36 to -33)
uncD (atp synthetase)	--GAGG(-11 to -8)	-GAUC-(-54 to -51)
uncC (atp synthetase)	-GGAGG(-10 to -6)	--AUCC(-18 to -15)

it, can control the interchange of base pairing.

Prediction of Translation-initiation Efficient Genes from E.coli We have scanned the EMBL data base for all the other *E.coli* genes for which the sequence of the −55 to −1 region is available. Each of these regions was searched for a well-defined SD site and a proposed TP site. These sites were subjected to a stringent analysis to examine that all the bases of the sites were free from any possible intramolecular secondary structure. Only 15 out of the 541 mRNAs were found to satisfy the above criteria; and therefore, would be expected to have high translation-initiation efficiency. These mRNAs are enlisted in Table 6. It should be noted that since translation-initiation is only one of the several possible controls operating in gene expression, we do not intend to claim that the above genes will be necessarily highly expressed. However, it appears that a majority of these genes codes for functionally important enzymes involved in main metabolic pathways and processes such as ATP synthesis and synthesis of important precursors.

CONCLUSION

In *E.coli*, the highly expressed genes appear to be equipped with additional ribosome-binding sites. Various sites involved in the process of initiation are shown in Figures 2 and 3. The occurrence of TP sites on highly expressed genes appear to enhance the translational efficiency in the following ways : (i) because of the additional contribution of the TP site to the strength of binding to the ribosomes, a relatively stabler 30S initiation complex can be formed, leading to a higher rate of formation of the 70S initiation complex; (ii) the anti-TP site on 16S rRNA can act as a primary recognition site for the interaction between the 30S complex and the 50S subunit; and (iii) the TP site on mRNA, which is released upon the formation of the 70S complex, can provide a mechanism whereby the next available free 30S ribosome can bind to mRNA even before the active ribosome finishes translating the initiator region, consequently increasing the size of the polysome. Considering the facts that the ribosome protects the mRNA from nucleolytic attack (39) and that the 5′ noncoding region appears to be a determinant of mRNA stability (40), the early binding of mRNA to the 30S complex through the TP site could enhance the mRNA stability leading to an improved translational efficiency. It is tempting to speculate that the TP-site and its proper manipulation could serve as a tool to maximise gene expression in terms of the efficiency of the protein synthesising machinery. The extension of this proposal to eukaryotic systems will be published separately.

ACKNOWLEDGEMENT

We thank EMBL for providing us the nucleic acid sequence database.

*To whom correspondence should be addressed

REFERENCES

1. Gold, L., Pribnow, D., Schneider, T., Singer, B.S. & Stormo, G. (1981) Annu. Rev. Microbiol. **35**, 365−403.
2. Kozak, M. (1983) Microbiol. Rev. **47**, 1−45.
3. Gren, E.J. (1984) Biochimie **66**, 1−29.
4. Steitz, J.A. (1980) In Chambliss, G., Craven, G.R., Davies, J., Davies, K., Kahan, L. and Nomura, M. (eds), Ribosomes : Structure, Function and Genetics, University Park, Baltimore, pp. 479−495.
5. Gouy, M. & Gautier, C. (1982) Nucleic Acids Res. **10**,7055−7074.
6. Ikemura, T. (1985) Mol. Biol. Evol. **2**, 13−34.
7. Liljenstrom, H. & Heijne, G.V. (1987) J. Theor. Biol. **124**, 43−45.
8. Ryoji, M., Berland, R. & Kaji, A. (1981) Proc. Natl. Acad. Sci. USA **78**, 5973−5977.

9. Bergmann, J.E. & Lodish, H.F. (1979) J. Biol. Chem. **254**, 11927–11937.
10. Jay, E., Seth, A.K. & Jay, G. (1980) J. Biol. Chem. **255**, 3809–3812.
11. Mitra, U., Stringer, E.A. & Chaudhuri, A. (1982) Annu. Rev. Biochem. **51**, 869–900.
12. Stormo, G.D. (1986) In Reznikoff, W.S. and Gold, L. (eds), Maximising Gene Expression, Butterworths, Massachusetts, pp. 195–224.
13. Shine, J. & Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA **71**, 1342–1346.
14. Steitz, J.A. & Jakes, K. (1975) Proc. Natl. Acad. Sci. USA **72**, 4734–4738.
15. Gheysen, P., Iserentant, D., Derom, C. & Fiers, W. (1982) Gene **17**, 55–63.
16. Hall, M.N., Gabay, J., Debarbouille, M. & Schuartz, M. (1982) Nature (London) **295**, 616–618.
17. Backendorf, C., Overbeek, G.P., Van Boom, J.H., Van der Marel, G., Veeneman, G. & Van Duin, J (1980) Eur. J. Biochem. **11**, 599–604.
18. Jansone, I., Berzin, V., Gribanov, V. & Gren, E.J. (1979) Nucleic Acids Res. **6**, 1747–1760.
19. Borisova, G.P., Volkova, T.M., Berzin, V., Rosenthal, G. & Gren, E.J. (1979) Nucleic Acids Res. **6**, 1761–1774.
20. Kastelein, R.A., Berkhout, B., Overbeek, G.P. & Van Duin, J. (1983) Gene **23**, 245–254.
21. Thompson, J.F. & Hearst, J.E. (1983) Cell **33**, 19–24.
22. Pelletier, J. & Sonenberg, N. (1987) Biochem. Cell Biol. **65**, 576–581.
23. Thanaraj, T.A., Kolaskar, A.S. & Pandit, M.W. (1988) J. Biomol. Struct. Dyn. **6**, 587–592.
24. Heus, H.A. & Van Knippenberg, P.H. (1988) J. Biomol. Struct. Dyn. **5**, 951–963.
25. Poldermans, B., Goosen, N. & Van Knippenberg, P.H. (1979) J. Biol. Chem. **254**, 9090–9094.
26. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Ceruthers, M.H., Neilson, T. & Turner, H.D. (1986) Proc. Natl. Acad. Sci. USA **83**, 9373–9377.
27. Curry, K.A. & Tomich, C.S.C. (1988) DNA **7**,173–179.
28. Varenne, S., Buc, J., Lloubes, R. & Lazduriski, C. (1984) J. Mol. Biol. **180**, 549–576.
29. Baga, M., Goransson, M., Normark, S. & Uhlin, B.E. (1988) Cell **52**, 197–206.
30. McCarthy, J.E.G., Schairer, H.U. & Sebald, W. (1985) EMBO J. **4**, 519–526.
31. Shpaer, E.G. (1986) J. Mol. Biol. **188**, 555–564.
32. Burma, D.P., Nag, B. & Tewari, D.S. (1983) Proc. Natl. Acad. Sci. USA **80**, 4875–4878.
33. Herr, W. & Noller, F. (1978) Biochemistry **17**, 307–315.
34. Herr, W. & Noller, H.F. (1979) J. Mol. Biol. **130**, 421–432.
35. Herr, W., Chapman, N.M. & Noller, H.F. (1979) J. Mol. Biol. **130**, 433–449.
36. Douthwaite, S., Christensen, A. & Garrett, R.A. (1983) J. Mol. Biol. **169**, 249–279.
37. Noller, H.F. (1984) Annu. Rev. Biochem. **53**, 119–162.
38. Van Duin, J., Kurland, C.G., Dondon, J., Grunberg-Manago, M., Branlant, C., & Ebel, J.P. (1976) FEBS Lett. **62**, 111–114.
39. Schneider, E., Blundell, M. & Kennel, D. (1978) Mol. Gen. Genet. **160**, 121–129.
40. Melefors, O. & Gabain, A.V. (1988) Cell **52**,893–901.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.