J. Biochem. 129, 851–860 (2001)

Codon and Base Biases after the Initiation Codon of the Open Reading Frames in the *Escherichia coli* Genome and Their Influence on the Translation Efficiency[1]

Toru Sato,[*] Mahito Terabe,[*] Hidemi Watanabe,[†] Takashi Gojobori,[‡] Chie Hori-Takemoto,[*] and Kin-ichiro Miura[*,2]

[*]Institute for Biomolecular Science, Gakushuin University, 1-5-1 Mejiro, Toshima-ku, Tokyo 171-8588; [†]Human Genome Research Group, Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045; and [‡]Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima 411-8540

Received February 1, 2001; accepted February 28, 2001

Nucleotide sequences around the boundaries of all open reading frames in the *Escherichia coli* whole genome were analyzed. Characteristic base biases were observed after the initiation codon and before the termination codon. We examined the effect of the base sequence after the initiation codon on the translation efficiency, by introducing mutations after the initiation codon of the *E. coli* dihydrofolate reductase (DHFR) gene, considering codon and base biases, and using *in vitro* and *in vivo* translation systems. In both assay systems, the two most frequent second codons, AAA and AAU, enhanced the translation efficiency compared with the wild type, whereas the effects of lower frequency codons were not significant. Experiments using 16S rRNA variants with mutations in the putative complementary sequence to the region downstream of the initiation codon showed that the translation efficiency of none of the DHFR mutants was affected. These results demonstrate that the statistically most frequent sequences for the second codon enhance translation efficiency, and this effect seems to be independent of base pairing between mRNA and 16S rRNA.

Key words: codon and base biases, *Escherichia coli*, 16S rRNA, second codon, translation efficiency.

Protein synthesis is a complex, continuous process comprising the main stages of initiation, elongation, and termination. The initiation phase is believed mainly to control translation in bacterial cells (*1*). A well-characterized element that influences the initiation step is the Shine-Dalgarno (SD) sequence, purine-rich nucleotides in the upstream of an open reading frame (ORF) (*2*). The SD sequence interacts with the complementary sequence in the 16S rRNA by base-pairing, which promotes the binding of mRNA to the 30S ribosomal subunit to form a pre-initiation complex. The importance of the base pairing between the SD sequence and 16S rRNA during the translational initiation step has been established by a series of biochemical experiments (*2, 3*). In the case of mRNA without an SD sequence, a ribosomal protein, S1, and initiation factor 3 have been shown to be helpful for the formation of the complex between the mRNA and 30S subunit (*4*). Sometimes the secondary structure of the mRNA around the initiation codon is involved in translational regulation by its own product, such as a ribosomal protein S15 or bacteriophage MS2 A protein (*5–7*). However, it has not been clarified why translation in almost all cells starts from the sequence AUG, the initiation codon, and how an initiator tRNA finds the correct position for initiation.

Over the past several years, whole genomes of various organisms have been sequenced completely. It is possible that a clue to the above questions might exist in the genome sequences. Statistical studies of the genome sequences of yeast and some bacteria have shown characteristic distributions of nucleotide bases around the initiation and termination codons (*8, 9*). These regions of mRNA are important for the interactions with the translational machinery, and it is expected that the nucleotide sequences in these regions might affect the efficiency of translational initiation and termination. In the case of *Escherichia coli*, adenine and guanine are highly biased in the region 6 to 12 bases upstream of the initiation codon. This region corresponds to the SD sequence (*2*). At the base following the termination codon, thymidine is highly biased. It has been suggested that the identity of this base determines the efficiency of translational termination in *E. coli* (*10, 11*). There also exist base biases after the initiation codon and before the termination codon. Thus it can be hypothesized that the nucleotide sequences in these regions also affect the efficiency of translation in the initiation and termination steps.

[1] This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan and the Japan Society for the Promotion of Science ("Research for the Future" Program, Japan.

[2] To whom correspondence should be addressed. Tel: +81-3-5992-1033, Fax: +81-3-5992-1034 E-mail: bio-dir@gakushuin.ac.jp

Abbreviations: DB, downstream box; DHFR, dihydrofolate reductase; HRP, horseradish peroxidase; L, 4-amino-2-(N^5-lysino)-1-(β-D-ribofuranosyl)pyrimidinium ("lysidine"); ORF, open reading frame; SD sequence, Shine-Dalgarno sequence; TBS, Tris-buffered saline.

We present here the relationship between the biases of nucleotide sequence and the efficiency of bacterial translation, using an *E. coli* translation system with the dihydrofolate reductase (DHFR) gene as a reporter. The translation efficiency as affected by sequences corresponding to the statistical study was assessed in *in vitro* and *in vivo* translation systems. Statistically favored sequences enhanced translation efficiency. The effects on translation efficiency observed in this study are discussed with reference to the results of *in vivo* assays using 16S rRNA variants.

MATERIALS AND METHODS

Genome Sequence Data—Complete genome sequence data for *E. coli* were obtained from the web site, http://www.genetics.wisc.edu (the DDBJ/GenBank/EMBL accession number U00096).

Estimation of Codon Biases—To evaluate the differences between the observed and expected codon frequencies statistically, the *G*-test was used as in the previous reports (*8, 9*).

Materials—Anti-rabbit HRP conjugated IgG was purchased from Gibco BRL and Ni-NTA Conjugates were from QIAGEN. The antibody against *E. coli* peptide release factor 1 was prepared by Sawaday Technology.

Primers—P1: 5′-GAAAAGGAGGAACTTCCATGGTCAGTCTGCTAGCGGCGTTAGCGGTAG-3′; P2: 5′-GGATCCCAATTGGTTAACAGATCTAAGCTTAACTAAC-3′; P3: 5′-GTGGTGCTCGAGTTTGCCATCTGAATCCATCAGG-3′; P4: 5′-GTATACACTCCGCCAGCGCTGAGGTC-3′; P5: 5′-GACCTCAGCGCTGGCGGAGTGTATAC-3′; P6: 5′-CCATTGCTGCAGCGGCGTAGAGGATCGAGATC-3′; P7: 5′-TGGTTGAGTACTTACAGGGCGCGTCCCATTC-3′; 16S1: 5′-CGCTTACCACTTTGTGCAGCATGACTGGGGTG-3′; 16S2: 5′-CGCTTACCACTTTGTGGTTCATGACTGGGGTG-3′; 16S3: 5′-CGCTTACCACTTTACTATTCATGACTGGGGTG-3′; 16S4: 5′-CGCTTACCACTTTGTGAATCATGACTGGGGTG-3′.

Construction of Plasmids—Plasmid pNCHE, used in the *in vitro* translation assay, was derived from pTHL5-12, into which was inserted a T7 phage promoter and an *E. coli dhfr* gene as described previously (*12*). A *Nco*I site and a *Nhe*I site were introduced around the initiation codon of the *dhfr* gene by the P1 primer. All series of mutants were made by insertion of annealed oligonucleotides at the *Nco*I-*Nhe*I sites (underlined sequences).

For the *in vivo* assay, the His-tagged *dhfr* gene was constructed as follows. Fragments of the *dhfr* gene were amplified by PCR using pNCHE as a template with two primers, P2 for the upstream and P3 for the terminal of the *dhfr* gene, and were then digested with *Mun*I and *Xho*I for insertion into *Eco*RI–*Xho*I digested pET21(+) (Novagen). For co-transformation with 16S rRNA mutants, the His-tagged *dhfr* gene fragments were inserted into pACYC177-Δ*Nhe*I, which lacked the *Nhe*I site of pACYC177, by the PCR primers P4 and P5. The His-tagged *dhfr* gene fragments were amplified by PCR using P6 primers for the upstream region of the T7 promoter, and P7 for the downstream region of the T7 terminator. These were ligated at the *Pst*I and *Sca*I sites to prepare pACYC177-Δ*Nhe*I (pAC-NCHE).

Mutagenesis of Plasmid-Encoded 16S rRNA—Plasmid pKK1192U derived from pKK3535 (*13*), which confers resistance to spectinomycin (*14*), was a kind gift from Drs. Michael O'Conner and Albert E. Dahlberg, Brown University. All mutations were introduced into the plasmid pKK-1192U by PCR primers, 16S1, 16S2, 16S3, and 16S4 using a QuikChange™ site-directed mutagenesis kit (Stratagene).

In Vitro Transcription and In Vitro Translation Reaction—pNCHE or its derivatives were linearized by digestion with *Pst*I and transcribed by T7 RNA polymerase. All transcripts were purified by application to MicroSpin S-400 HR Columns (Pharmacia). The amounts of RNA transcripts collected by ethanol precipitation were calculated from the absorbance at 260 nm.

In vitro translation assays in the *E. coli* S30 Extract System for Linear Templates Kit (Promega) were carried out according to the manufacturer's protocol as follows. The reaction mixture contained 5 μl of an amino acid mixture minus methionine, 20 μl of S30 Premix without amino acids, 1 μl of [^{35}S]methionine (15 mCi/ml; Amersham), 15 μl of the S30 Extract, and 1.5 μg of mRNA was incubated at 37°C for 2 h. The products were collected by trichloroacetic acid precipitation and filtering under vacuum on Whatman GF/A glass fiber filters. The total radioactivity of the translation products retained in the glass filters was measured by a liquid scintillation counter. To confirm the length of the translation products, the proteins in the reaction mixture were precipitated with acetone, separated by 15% SDS-PAGE, and exposed to X-ray film (Fujifilm).

In Vivo Protein Synthesis Analysis—For *in vivo* protein synthesis analysis of His-tagged DHFR, BL21(DE3) cells were transformed with a pAC-NCHE series and grown on LB plates containing kanamycin (20 μg/ml). Single colonies were incubated overnight in 1 ml of LB media with kanamycin. The grown cells were diluted 1:100 in 3 ml of fresh medium, and allowed to grow to an OD_{600} of 1.0. Then IPTG was added to 1 mM, and the cells were incubated for 2 h. Cells were then harvested from 500 μl of each culture by centrifugation, lysed in sample buffer (0.01 M sodium phosphate pH 7.2, 1% SDS, 1% β-mercaptoethanol, 6 M urea) and heated. Proteins were separated by 15% SDS-PAGE and transferred to nitrocellulose membranes (Schleicher & Schuell). The membranes were blocked for 1 h in TBS with 3% BSA and separated in order to detect of DHFR and peptide release factor 1 (RF1) as a standard marker. DHFR and RF1 were reacted with Ni-NTA Conjugates (1:5,000 dilution) and RF1 antibody (1:5,000 dilution), respectively. The membrane for RF1 was further incubated for 1 h with anti-rabbit HRP conjugated IgG (1:8,000 dilution). For detection of these proteins, the membranes were incubated with SuperSignal Substrate Western Blotting (Pierce) and exposed to X-ray film (Fujifilm). The signals on the films were quantified by the software NIH image and the ratio of the amount of expressed proteins was calculated.

Translation Analysis with Plasmid-Encoded 16S rRNA Mutant—Single colonies of BL21(DE3) co-transformed with pAC-NCHE and pKK1192U series were cultured in 1 ml of LB media containing ampicillin (50 μg/ml), kanamycin (20 μg/ml), and spectinomycin (50 μg/ml), and incubated for overnight. The cells were diluted 1:100 in 5 ml of fresh medium, and allowed to grow to an OD_{600} of 1.0, after

which IPTG was added to 1.0 mM. After incubation for 3 h, 1 ml aliquots from each culture were removed. The cells were harvested and the following procedures were the same as described above.

RESULTS

Codon Biases in the 5′- and 3′-Terminal Regions in E. coli ORFs—DNA sequences of 4,285 ORFs were obtained from an *E. coli* genome database. The appearance frequencies of 61 codons in 4,285 ORFs were counted for codon numbers 1 to 20 and −1 to −20. These frequencies for codon numbers 1 to 5 are shown in Table I. It is remarkable that CTG, which is the most frequent codon in the coding region of the *E. coli* genome (5.28% of total codons), is rare at the second codon (1.98%). In contrast, AAA, which accounts of 3.37% of total codons is frequent at the second codon (10.1%). Figure 1 shows that AAA appears frequently compared with its average appearance frequency at both termini of the ORFs. On the other hand, CTG appears less frequently than its

TABLE I. **Codon bias in the 5′-terminal region of the *E. coli* ORFs.**

Codon 1			Codon 2			Codon 3			Codon 4			Codon 5		
Codon(AA)	Freq	*G*-term	Codon(AA)	Freq	*G*-term	Codon(AA)	Freq	*G*-term	Codon(AA)	Freq	*G*-term	Codon(AA)	Freq	*G*-term
ATG(M)	3540	23973.83	AAA(K)	433	950.74	AAA(K)	306	459.43	AAA(K)	258	299.32	AAA(K)	222	190.83
GTG(V)	613	2068.27	AAT(N)	188	339.21	GAA(E)	170	0.84	TTA(L)	182	405.55	ATT(I)	197	162.39
TTG(L)	130	206.54	AGT(S)	158	452.12	AAG(K)	161	414.90	ATT(I)	180	115.89	TTT(F)	154	146.02
ATT(I)	1	-9.74	AGC(S)	151	236.19	ATT(I)	159	62.92	TTT(F)	147	125.70	TTA(L)	136	223.80
CTG(L)	1	-10.84	ATT(I)	136	11.32	AAT(N)	138	163.66	ATC(I)	132	52.90	GAA(E)	132	-66.14
AAA(K)	0	-	ATG(M)	134	30.04	CAG(Q)	137	27.60	CAA(Q)	131	179.86	CTG(L)	130	-143.91
AAC(N)	0	-	GCT(A)	130	177.81	ATG(M)	136	34.52	GAA(E)	129	-70.57	AAT(N)	127	129.52
AAG(K)	0	-	GAA(E)	126	-74.86	GAT(D)	131	-13.95	AAT(N)	127	129.52	ATC(I)	125	36.47
AAT(N)	0	-	AAC(N)	125	73.63	AAC(N)	128	81.47	CTG(L)	113	-156.77	CAA(Q)	117	134.19
ACA(T)	0	-	GCA(A)	123	86.30	CAA(Q)	118	137.35	AAC(N)	103	20.79	GAT(D)	108	-53.20
ACC(T)	0	-	AAG(K)	113	211.20	TTA(L)	112	140.82	ATG(M)	100	-36.12	ACC(T)	106	11.20
ACG(T)	0	-	TCT(S)	104	218.74	TTT(F)	112	34.86	CGT(R)	94	8.73	ATG(M)	101	-34.47
ACT(T)	0	-	ACA(T)	102	246.08	ACA(T)	108	272.90	GCA(A)	92	11.12	GTT(V)	96	38.53
AGA(R)	0	-	CAA(Q)	99	80.47	ATC(I)	92	-29.56	ACA(T)	89	190.45	ACA(T)	88	186.32
AGC(S)	0	-	GAT(D)	99	-66.00	ACG(T)	83	48.26	GAT(D)	86	-81.55	CGC(R)	86	-16.19
AGG(R)	0	-	ACC(T)	95	-10.78	TAT(Y)	83	29.41	GTT(V)	84	11.28	AAC(N)	85	-15.49
AGT(S)	0	-	ACT(T)	90	152.67	ACC(T)	82	-33.44	ATA(I)	81	236.51	ACT(T)	85	134.47
ATA(I)	0	-	TTT(F)	90	-11.35	CTG(L)	76	-165.73	TAT(Y)	80	22.46	CGT(R)	85	-9.21
ATC(I)	0	-	CTG(L)	85	-166.33	ACT(T)	74	96.56	AAG(K)	77	84.84	TAT(Y)	82	27.07
CAA(Q)	0	-	ACG(T)	81	43.14	GTT(V)	74	-8.82	CTT(L)	77	74.86	CTT(L)	77	74.86
CAC(H)	0	-	GCG(A)	79	-95.53	ATA(I)	68	174.76	CAG(Q)	74	-76.25	GCA(A)	76	-19.86
CAG(Q)	0	-	TCA(S)	77	140.71	CGT(R)	65	-41.92	TCA(S)	71	118.23	CAG(Q)	73	-77.21
CAT(H)	0	-	CGT(R)	76	-25.25	GAG(E)	64	-23.11	ACC(T)	70	-50.70	GTA(V)	73	64.84
CCA(P)	0	-	ATC(I)	71	-59.60	GCA(A)	64	-38.72	TTC(F)	70	-2.34	ACG(T)	72	21.39
CCC(P)	0	-	CAG(Q)	69	-80.75	GCG(A)	62	-105.02	AGC(S)	66	-6.01	AAG(K)	70	63.78
CCG(P)	0	-	TTA(L)	63	6.71	TCA(S)	62	86.43	CGC(R)	66	-47.36	ATA(I)	66	165.68
CCT(P)	0	-	CGC(R)	57	-57.62	AGT(S)	60	55.50	CTC(L)	66	43.07	TCA(S)	63	89.84
CGA(R)	0	-	TCC(S)	54	40.55	CAT(H)	58	4.94	TTG(L)	65	13.16	TGG(W)	63	-4.79
CGC(R)	0	-	ATA(I)	47	86.07	AGC(S)	57	-21.90	TCT(S)	63	69.35	GCT(A)	62	-7.01
CGG(R)	0	-	CCA(P)	47	24.41	CGC(R)	55	-59.52	ACT(T)	61	56.02	CAT(H)	61	11.35
CGT(R)	0	-	CCT(P)	47	41.67	CTT(L)	55	16.46	CCA(P)	58	54.52	TTG(L)	61	4.60
CTA(L)	0	-	TTC(F)	47	-39.02	GAC(D)	55	-43.89	GTA(V)	57	22.42	AGC(S)	59	-18.60
CTC(L)	0	-	CAT(H)	46	-17.41	AGA(R)	54	191.83	ACG(T)	56	-11.51	GCG(A)	59	-105.79
CTT(L)	0	-	CTT(L)	46	-2.67	TCT(S)	52	37.28	CCG(P)	54	-66.20	GTC(V)	58	-14.26
GAA(E)	0	-	TCG(S)	45	14.15	GCT(A)	50	-27.16	GTC(V)	51	-25.66	CTC(L)	57	20.48
GAC(D)	0	-	TTG(L)	45	-23.98	GTA(V)	49	4.46	GGA(G)	50	37.47	TCT(S)	57	51.33
GAG(E)	0	-	CCG(P)	44	-71.97	TAC(Y)	49	-6.87	AGT(S)	49	25.48	TTC(F)	57	-25.33
GAT(D)	0	-	GAG(E)	43	-49.72	CGA(R)	48	109.61	TAC(Y)	48	-8.71	CAC(H)	56	32.66
GCA(A)	0	-	CGA(R)	38	69.02	CAC(H)	47	10.94	TGG(W)	48	-29.76	GTG(V)	56	-79.07
GCC(A)	0	-	GGT(G)	38	-78.22	TGG(W)	47	-31.12	GGG(G)	47	-93.72	GAC(D)	55	-43.89
GCG(A)	0	-	GTT(V)	38	-55.18	TTC(F)	47	-39.02	GAC(D)	46	-53.15	GCC(A)	49	-78.93
GCT(A)	0	-	GGA(G)	37	5.44	GGT(G)	46	-77.11	TCC(S)	46	19.79	TCC(S)	45	17.38
GGA(G)	0	-	AGA(R)	36	98.69	GTC(V)	44	-35.13	GCC(A)	44	-80.34	TAC(Y)	43	-17.26
GGC(G)	0	-	GTA(V)	36	-18.92	TTG(L)	44	-25.43	CGA(R)	42	84.69	AGA(R)	42	128.09
GGG(G)	0	-	GAC(D)	35	-59.57	GGA(G)	41	14.45	TCG(S)	42	7.41	AGT(S)	42	8.89
GGT(G)	0	-	GCC(A)	35	-79.93	GTG(V)	40	-83.40	CAT(H)	41	-24.95	CCA(P)	42	12.37
GTA(V)	0	-	TAT(Y)	30	-50.43	GGC(G)	39	-92.32	CCT(P)	41	25.15	CCT(P)	42	27.79
GTC(V)	0	-	CTC(L)	28	-29.75	CTC(L)	38	-17.16	GCT(A)	41	-38.55	CCC(P)	40	42.15
GTT(V)	0	-	CAC(H)	27	-23.65	GCC(A)	38	-80.53	GAG(E)	39	-52.71	CCG(P)	40	-73.05
TAC(Y)	0	-	CTA(L)	27	25.74	CCA(P)	34	-4.36	GCG(A)	38	-101.57	GGC(G)	39	-92.32
TAT(Y)	0	-	GGC(G)	27	-83.77	TCC(S)	34	-5.93	GTG(V)	38	-83.13	GGT(G)	35	-77.80
TCA(S)	0	-	GTC(V)	24	-48.26	CCT(P)	33	5.92	AGA(R)	37	103.46	TCG(S)	32	-11.76
TCC(S)	0	-	TAC(Y)	24	-37.63	CCG(P)	27	-70.53	CCC(P)	37	33.22	TGC(C)	28	0.21
TCG(S)	0	-	CGG(R)	23	-0.69	TCG(S)	26	-20.35	CAC(H)	34	-14.10	GAG(E)	27	-56.35
TCT(S)	0	-	TGG(W)	22	-47.97	AGG(R)	25	77.17	CTA(L)	30	34.92	CGA(R)	25	24.47
TGC(C)	0	-	AGC(R)	21	57.50	CGG(R)	25	3.42	GGT(G)	27	-74.03	CTA(L)	25	19.99
TGG(W)	0	-	GTG(V)	19	-67.90	CCC(P)	24	0.77	AGG(R)	25	77.17	TGT(C)	25	5.79
TGT(C)	0	-	GGG(G)	18	-35.06	CTA(L)	23	14.55	GGG(G)	23	-33.53	GGA(G)	22	-19.64
TTA(L)	0	-	CCC(P)	17	-11.18	TGC(C)	23	-8.87	TGT(C)	23	1.49	GGG(G)	20	-34.74
TTC(F)	0	-	TGT(C)	10	-16.01	GGG(G)	17	-35.06	TGC(C)	20	-13.31	AGG(R)	16	35.11
TTT(F)	0	-	TGC(C)	10	-20.52	TGT(C)	16	-10.58	CGG(R)	19	-7.83	CGG(R)	13	-15.22
G-value		26228.06			2323.38			1557.48			1456.39			1132.36

Appearance frequencies and *G*-terms of codons are shown for codon numbers 1 to 5. *G*-term values are estimated at 0 for frequency 0 in the calculation of *G*-value.

average in these regions.

In order to evaluate statistically the difference between the observed and average appearance frequencies of codons, the *G*-test (otherwise called the likelihood chi-square test) was applied as previously reported (*8, 9*). As the traditionally used chi-square test gives positive values as the deviation of an observed value from the expected one, it is inconvenient to identify whether the observed frequency is greater or smaller than expected. The *G*-value and the chi-square value are asymptotically the same (Eq. 1).

$$G = 2\Sigma O \ln (O/E) \cong \chi^2 = \Sigma (O-E)^2/E, \quad (1)$$

where O and E are the observed and expected values, respectively. In this study, the appearance frequencies described above were used as the observed codon frequencies (N_{obs}). The expected codon frequencies (N_{exp}) were estimated as follows. The fractions of each codon among 1,358,928 codons, that is a total of 4,285 ORFs except termination codons, were calculated. Using these fractions, the average appearance frequency of each codon was calculated and used as the expected frequency (N_{exp}). In this procedure, N_{exp} for each codon is always the same value, independent of the codon number. In Table I, all codons are lined up in order of appearance frequency (N_{obs}) for codon numbers 1 to 5. *G*-term values [the values of $2 \times N_{obs} \ln (N_{obs}/N_{exp})$] of each codon are also given in Table I. At the first codon, the *G*-term is extremely high, because ATG is used exclusively as the initiation codon in the *E. coli* genome. As the reason for the high *G*-term is clear, the codon bias at the first codon is not dealt with in this article. The *G*-plot for codon numbers 2 to 20 and –1 to –20 is shown in Fig. 2A. In comparison with the middle region, high codon biases are observed in both ends of the ORFs. The chi-square test gave the same result as the *G*-test (*data not shown*). At codon numbers 2, 3, 4, and –1, the codon biases are also very strong. Characteristic base distributions were observed after the initiation codon and before the termination codon of *E. coli* ORFs. This suggests that codon biases in this region have some physiological

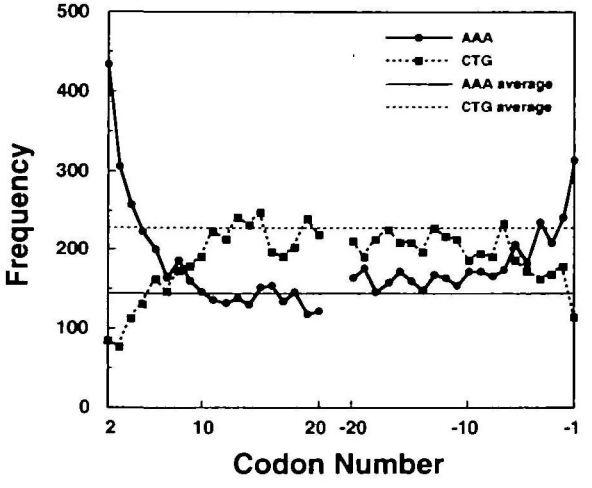

Fig. 1. **Appearance frequencies of codons AAA and CTG in two terminal regions of *E. coli* ORFs.** The appearance frequencies of codons AAA and CTG in 4,285 *E. coli* ORFs are indicated for codon numbers 2 to 20 and –1 to –20. The average frequencies of these codons are also shown.

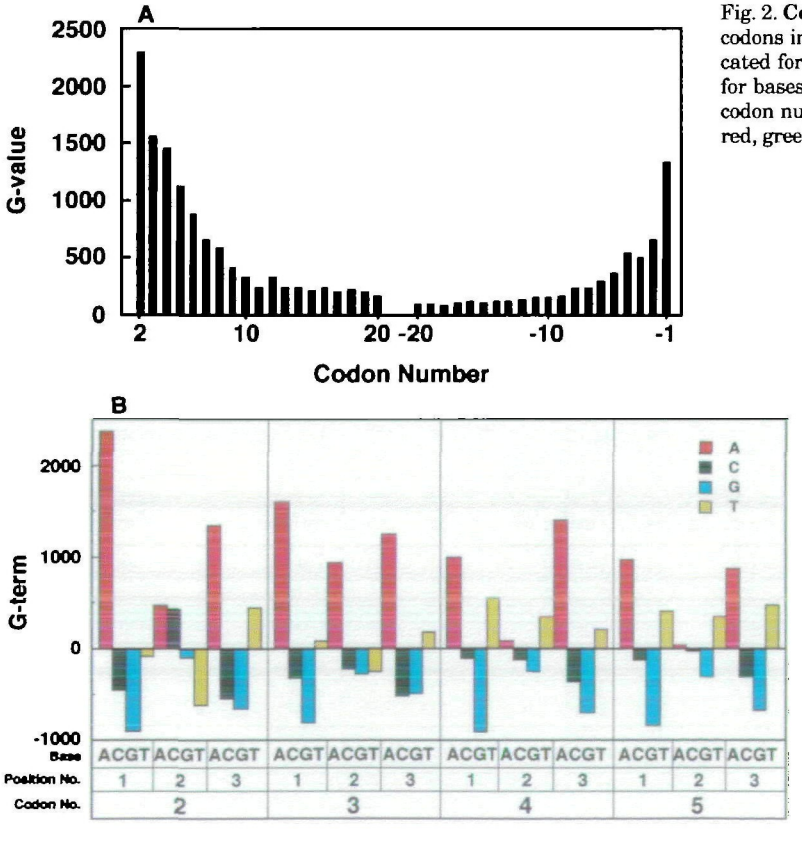

Fig. 2. **Codon and base biases in *E. coli* ORFs.** (A) *G*-values for codons in two terminal regions of *E. coli* ORFs. *G*-values are indicated for codon numbers 2 to 20 and –1 to –20. (B) *G*-term values for bases. *G*-term values for three base positions are indicated for codon numbers 2 to 5. Values for A, C, G, and T are represented in red, green, blue and yellow, respectively.

meaning in terms of translation efficiency.

Effects of Statistically Frequent Second Codon Sequences on Translation Efficiency—To investigate whether any relationship exists between codon biases after the initiation codon and translation efficiency, an *in vitro* translation study was performed using the *E. coli dhfr* gene as a reporter. Mutations were introduced into the pNCHE series resulting from pTHL5-12 possessing the T7 promoter upstream of the *dhfr* gene (*12*) as described in "MATERIALS AND METHODS" (Fig. 3). Figure 4A shows the second codon sequences of the DHFR mutants used here and the frequencies of each codon exhibited in Table I. AAA (corresponding to m3) and AAU (m1) are the two most frequent codons at the second codon, accounting for 10.1 and 4.4% of 4,285 ORFs, respectively. As the least frequent codons, UGC (m2) and GUC (m4) were selected. The most common codon in all ORFs in the *E. coli* genome, CUG (m5), and the rarest codon, AGG (m6), were also tested. The wild type sequence is derived from the original *dhfr* gene. The translation efficiency of these mRNAs was estimated using an *E. coli* S30 extract because it is necessary to investigate only the translation process. Figure 4B shows the incorporation of [^{35}S]methionine into the DHFR mutants as measured by liquid scintillation counting. The values were calculated relative to a wild type value of 1.0 after subtraction of the amount of incorporated [^{35}S]methionine with minus mRNA as the background. The highly frequent sequence mutants, m1 and m3, enhanced the translation efficiency. However, the values of the low frequency sequence mutants, m2 and m4, were approximately same as the wild type value. The value of mutant m6, the least frequent codon, was also same as the wild type. Somewhat interestingly, mutant m5, possessing a CUG codon in the second position, which is the most frequent codon in all ORFs and the lowest *G*-term in the second codon, showed approximately half the translation efficiency of the wild type. These results clearly show that the statisticaly more frequent sequences at the second codon of *E. coli* correlate with high translation efficiency, while low frequency sequences do not lead to low translation efficiency. These differences in translation efficiency were also demonstrated by autoradiography (Fig. 4C). The tendency was observed with a short reaction time (data not shown).

Effects of Various Second Codons on Translation Efficiency—To consider the enhancement of translation produced by the highest frequency sequences, we examined the translation efficiency of various second codon sequences. Figure 5 shows the relative efficiency of translation of DHFR mutants at the second codon in order of appearance frequency (Freq in Table I). The results for AAA, AAU, and AGU clearly show that there is a significant correlation between high *G*-term values and the enhancement of translation efficiency. There are some factors that might affect translation efficiency; for example, the nature of the amino acid, the affinity of aminoacyl-tRNA for the A-site in the ribosome, and the population of the corresponding tRNA (*15, 16*). If high translation efficiency is dependent on the nature of the tRNA, then synonymous codons should also enhance the translation efficiency. Although AAG, like AAA, is a lysine codon and is recognized by the same $tRNA^{Lys}$, the translation efficiency of the AAG mutant is not very increased. A similar relationship can be observed for the asparagine codons AAU and AAC. Furthermore, both AUU and AUA are isoleucine codons recognized by a major $tRNA_{GAU}$ and a minor $tRNA_{LAU}$ respectively (*17, 18*). But there is a large difference in the translation efficiency between the AUU and AUA mutants that is quite the reverse of the tRNA populations. The length of the translation product of each mutant was confirmed by autoradiog-

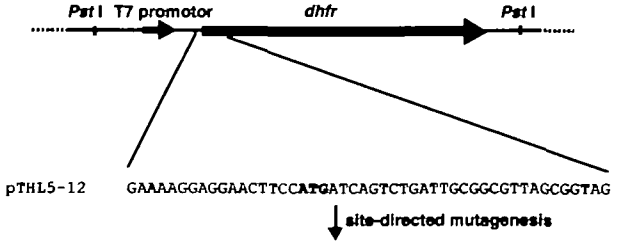

Fig. 3. **Schematic representation of plasmid construction.** The DNA sequence of pTHL5-12 around the initiation codon of the *E. coli dhfr* gene is represented. The initiation codon is displayed in bold type. *Nco*I and *Nhe*I sites are underlined.

A

		Freq	G-term
Wild Type:	...AUG AUC AGUCUG...	middle	low
m1:	...AUG AAU AGUCUG...	high	high
m2:	...AUG UGC AGUCUG...	low	low
m3:	...AUG AAA AGUCUG...	high	high
m4:	...AUG GUC AGUCUG...	low	low
m5:	...AUG CUG AGUCUG...	middle	low
m6:	...AUG AGG AGUCUG...	low	middle

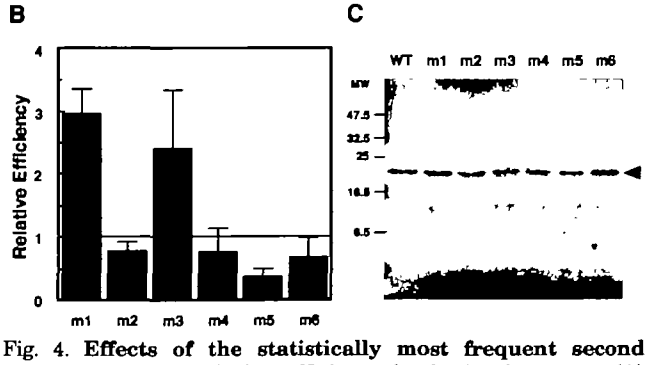

Fig. 4. **Effects of the statistically most frequent second codons on the translation efficiency in the *in vitro* assay.** (A) Sequences after the initiation codon of the *dhfr* mutants used in this study. Wild type sequence is derived from the original *dhfr* gene. The appearance frequency (N_{obs}) and *G*-term of the second codon are indicated in line after each sequence. (B) Relative translation efficiency of DHFR mutants calculated from the radioactivity of [^{35}S]-methionine incorporated into protein products determined by the filter assay. The error bars indicate standard deviation. To normalize the results, the wild type value is set at 1.0. These values represent the average of several independent experiments. (C) Autoradiogram showing [^{35}S]methionine labeled DHFR mutants produced in an *in vitro* translation system using equal amounts of mRNAs. Proteins were separated by SDS-PAGE (15%). Molecular mass markers (MW) are shown by size. The arrowhead indicates DHFR mutants.

raphy (data not shown). These results suggest that the observed enhancement of translation efficiency is caused by the mRNA sequence rather than the nature of the tRNA and amino acid.

Effects of the Third Codon on Translation Efficiency—We found that the most efficient sequence AUGAAU is complementary to a part of the 16S rRNA (nucleotide positions 1480–1485), and introduced a mutation, CAC, into the third codon to obtain nine nucleotides complementary to 5′-GUGAUUCAU-3′ in 16S rRNA (Fig. 6A). For comparison, the statistically most frequent sequences, AAA, or the codon yielding high translation efficiency, AAU, was substituted in the third codon, with the second codon fixed as AAU.

Figure 6B shows that the translation efficiencies of the AAU-1 and AAU-2 mutants were similar to that of m1, whose third codon was same as the wild type. Unexpectedly, the 16S rRNA complementary mutant, AAU-3, showed reduced translation efficiency compared with m1. Higher complementation between an ORF and 16S rRNA sequences reduces the translation efficiency. This suggests that this region right below the initiation codon interacts with the 16S rRNA.

Translation Efficiency of the Second Codon in the In Vivo Assay—The effects of the mutations described above on whole expression system were also examined *in vivo* in *E. coli*. Derivatives of pAC-NCHE contain *dhfr* mutant genes fused with 6 repeats of histidine residues (His-tag) at the carboxyl terminus, under the control of the T7 promoter and *lac* operator. These plasmids were introduced into *E. coli* strain BL21(DE3), and the expression levels of the mutant *dhfr* genes were analyzed. For the *in vivo* assay, we selected DHFR mutants that caused remarkable changes in the translation efficiency *in vitro*. DH1 and DH2 correspond to m1 and m3, respectively, mutants that enhanced the translation efficiency *in vitro*. DH4 and DH5 possess the codon synonymous to DH1 and DH2, respectively. DH3 corresponds to m5 (CUG), which showed a negative effect on translation in the *in vitro* assay (in Fig. 4B). DH6 possesses a sequence complementary to 16S rRNA, corresponding to AAU-3 (in Fig. 6). DH7 (AUU) differs in only one base from DH1, while the corresponding mutant shows a large difference in translation efficiency (Fig. 5). Figure 7B shows signals of western blots for DHFR and peptide release factor 1 (RF1) as a standard marker of translation activity in each transformed cell. The signals on the films were quantified by NIH image software and the ratios of the amount of expressed proteins to the wild type value as 1.0 were calculated (Fig. 7C). DH1 and DH2 enhanced the translation efficiency as well as in the *in vitro* assay. DH3 reduced the translation efficiency more notably than in the *in vitro* assay. DH4 did not enhance efficiency as well as DH1 despite the synonymous codon. The same relationship can be applied between DH5 and DH2. This suggests that the mRNA sequence in the second codon is important for translation efficiency. DH6 and DH7 reduced the translation efficiency somewhat. The results of the *in vivo* assay show that the effects of substitutions in the second codon on translation efficiency are similar to those in the *in vitro* assay.

Correlation between Sequences Downstream of the Initiation Codon and 16S rRNA—It has been proposed that there is a translational enhancer element located downstream of the initiation codon in several bacterial and phage mRNAs. This element is named the downstream box (DB) and is complementary to 16S rRNA nucleotides 1469–1483, the so-called antidownstream box (anti-DB) (*19–23*).

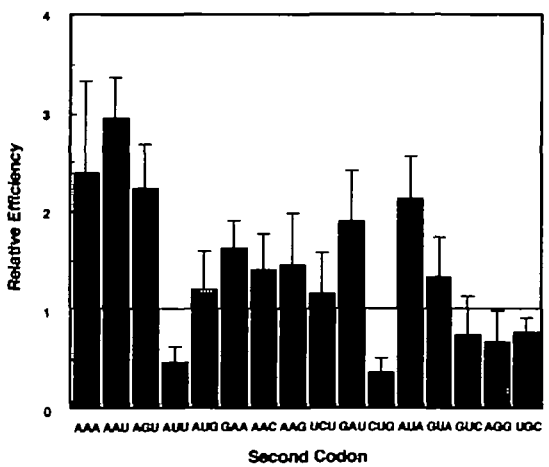

Fig. 5. **Effects of various second codons on the translation efficiency in the *in vitro* assay.** The protein products were labeled with [^{35}S]methionine, and the amounts were determined by their radioactivity. The error bars indicate standard deviation. To normalize the results, the wild type value is set at 1.0. The values represent the average of several independent experiments.

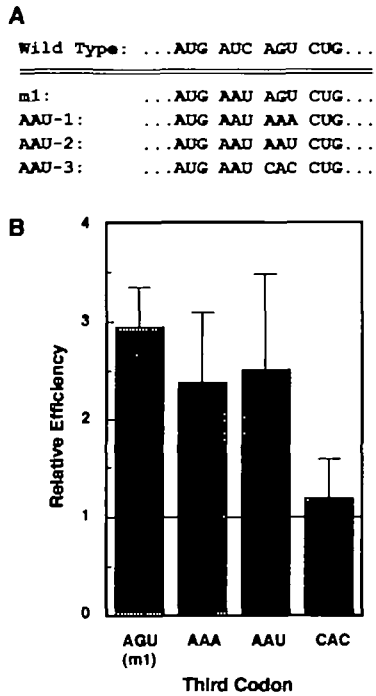

Fig. 6. **Effect of third codons on the translation efficiency in the *in vitro* assay.** (A) Sequences after the initiation codon of the *dhfr* mutants used in this study. In these mutants, the second codon was fixed as AAU. (B) Relative translation efficiency of DHFR mutants with different third codons. The protein products were labeled with [^{35}S]methionine, and the amounts were determined by their radioactivity. The error bars indicate standard deviation. To normalize the results, the wild type value is set at 1.0. These values represent the average of several independent experiments.

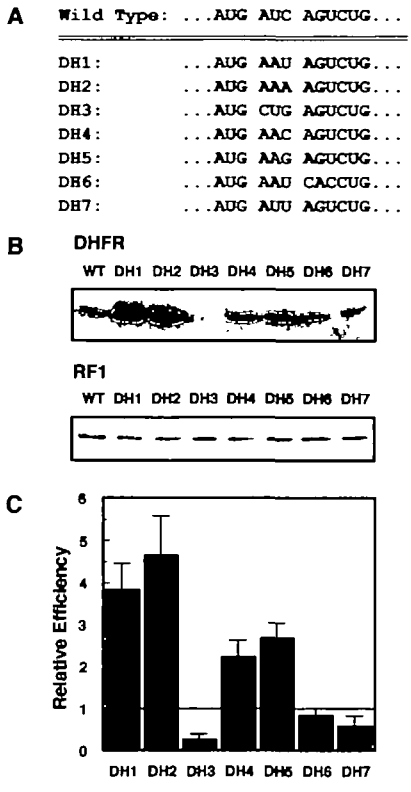

Fig. 7. **Effects of sequences downstream of the initiation codon on translation efficiency in the *in vivo* assay.** (A) Sequences after the initiation codon of the *dhfr* mutants used in this study. The sequences were derived from those of *dhfr* mutants with remarkable effects on translation efficiency in the *in vitro* assay. (B) Western blotting showing DHFR-His mutants expressed in *E. coli* strain BL21(DE3) using equal volumes of culture. Proteins were separated by SDS-PAGE (15%), and transferred to nitrocellulose membranes and then detected with Ni-NTA Conjugates (QIAGEN). Peptide release factor 1 (RF1) was used as an internal standard. (C) Relative translation efficiency calculated from the quantitative analysis of the signals observed in Fig. 7B by western blotting. The error bars indicate standard deviation. To normalize the results, the wild type value is set at 1.0. These values represent the average of several independent experiments.

We found that the nucleotide sequence AUGAAU, deduced in this study as the sequence showing the highest appearance frequency and the highest translation efficiency, forms part of the DB. Furthermore, the translation efficiency was affected when the complementarity to 16S rRNA was changed in this region. It is possible that the sequences downstream of the initiation codon of the mRNA interact directly with 16S rRNA and modulate translation efficiency. Although a recent report suggested that the enhancement of translation by DB did not involve an interaction between mRNA and rRNA (*24*), it has not yet been shown directly that the DB interacts with the anti-DB by base pairing. To clarify whether the translation efficiency is influenced by the uncertain interaction between DB and the putative anti-DB, we investigated expression levels of DHFR mutants using 16S rRNA variants with mutations introduced in the region of the putative anti-DB in the *in vivo* assay system. Figure 8 shows the sequences of the 16S rRNA variants used in this study. These variants and the wild type 16S rRNA were encoded in plasmid pKK1192U. The antibiotic spectinomycin inhibits the translation of genome-encoded ribosomes, but the expression of plasmid-encoded 16S rRNA containing a C1192 to U mutation supports growth in media containing spectinomycin (*14*). Therefore, in the presence of spectinomycin, all actively translating ribosomes result from plasmid-encoded 16S rRNA. We constructed a combination of DHFR mutants and 16S rRNA variants to examine the effects of base pairing between mRNA and 16S rRNA as described below. A 16S rRNA variant, 16S1, is complementary to the second codon of DH3 (Fig 8B). If there is any interaction between the second codon and 16S rRNA, the translation of DH3 should be more efficient, while that of DH1 might be reduced. Likewise, cells transformed by 16S2 or 16S4 are expected to exhibit similar results (Fig. 8, C and E). 16S3 is changed so as to be complementary to the third codon of DH1 (Fig. 8D). If there were any interaction, DH1 in 16S3 transformed cells would reduce the translation efficiency, similar to DH6 in cells transformed by wild type 16S rRNA.

The upper panels of Fig. 9 show western blots of the protein products for DHFR-His. The expression levels of RF1 were also determined and show that the translation efficiency was similar in all experiments (data not shown). The signals on the films were quantified and calculated (lower panels of Fig. 9). The numbering of each panel corresponds to that in Fig 8. Comparing Fig. 9A to Fig. 7C, there is no significant difference between the relative efficiency of translation using 16S rRNA encoded in plasmids and

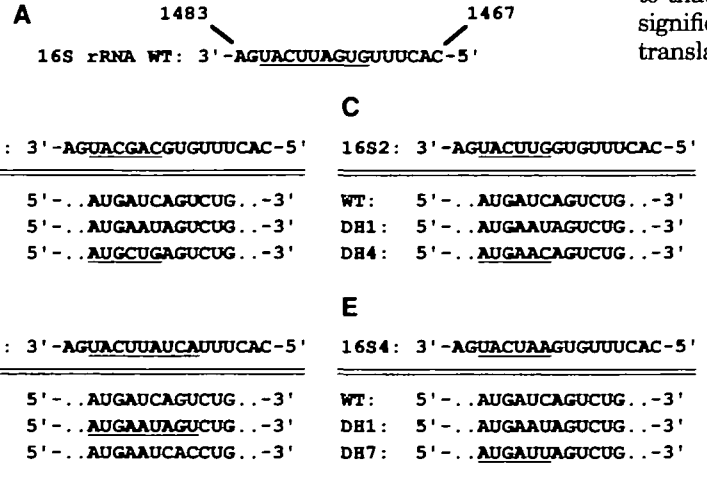

Fig. 8. **Sequences of 16S rRNA mutants in the anti-downstream box region and co-transfected DHFR mutants.** Each mutant and wild type 16S rRNA was encoded in plasmids pKK1192U, and transformed cells were cultured in media with the antibiotic spectinomycin. Underlined nucleotides in 16S rRNA and the *dhfr* gene are complementary sequences to each other.

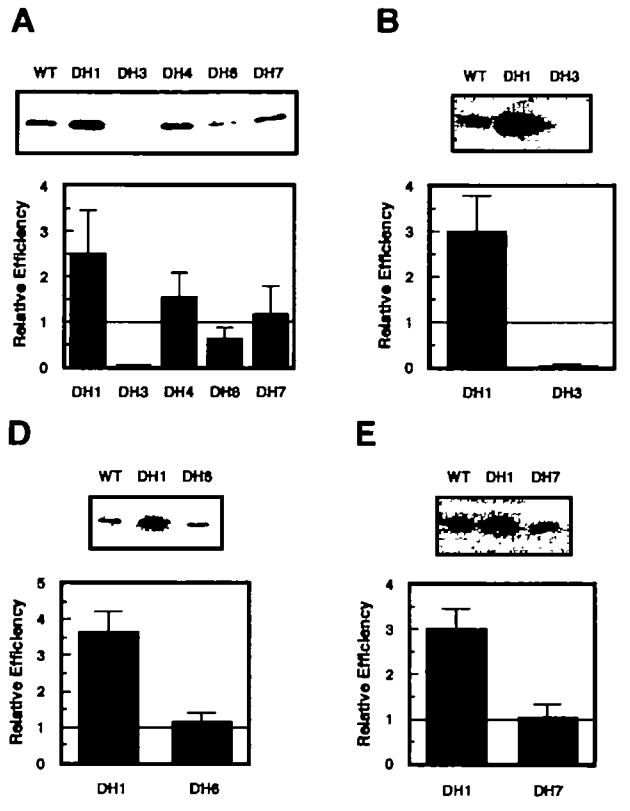

Fig. 9. **Influence of mutagenesis of 16S rRNA on the translation efficiency in co-transformed DHFR mutant cells.** The upper panels show the results of western blotting of DHRF mutants detected with Ni-NTA Conjugates. The relative efficiency was calculated by quantitative analysis of the western blotting signals. These values represent the average of four independent experiments. (A) Effect on the translation efficiency of DHFR mutants with wild type 16S rRNA encoded in the plasmid. The diagrams, (B), (C), (D), and (E) show the effects of the combinatorial co-transformation of 16S rRNA variants with *dhfr* gene mutants corresponding to Fig. 8. The values were calculated based on the amount of wild type DHFR with each 16S rRNA variants as 1.0. In all experiments, the internal control, RF1, showed the same expression level (data not shown).

genome-encoded 16S rRNA. The tendency of the change in translation efficiency for any DHFR mutant with 16S rRNA variants was same as that for wild type 16S rRNA. These results show that the sequences downstream of the initiation codon of mRNAs do not interact directly with the anti-DB region of 16S rRNA through the expected base pairs. However, the relative growth rates of cells transformed by 16S rRNA variants differ (data not shown). The data suggest that the anti-DB region does not form base pairs with DB in mRNA, while the 16S rRNA sequence around the anti-DB region influences protein synthesis and cell growth.

DISCUSSION

As shown in Table I and Figs. 1 and 2, the notable distribution of nucleotides observed in the two boundary regions of ORFs gives rise to codon biases. In these regions, amino acid biases are also observed (data not shown). The characteristic distributions observed in the three kingdoms of life (*8, 9*) allow to us to hypothesize that these distributions originated during evolution of each respective translation system. Here, we have investigated codon and base biases in the region downstream of the initiation codon. We considered that codon biases would have arisen phylogenetically due to advantages for protein synthesis. First, it must be asked whether the characteristic amino acid distribution causes the base and codon biases downstream of the initiation codon. In this case, the optimal codon should be chosen among synonymous codons as the second codon, because it has been suggested that codon usage and the cellular tRNA population could affect the production levels of proteins (*25, 26*). Actually, the highly expressed gene, *ompA*, a major outer membrane protein, has the codon, AAA at the second codon (*27*). The high translation efficiency is achieved by the high frequency codon. Although the most frequent codon, AAA, is the optimal codon for lysine, AAU and AGU are not the optimal codons for asparagine and serine, respectively (*15, 28, 29*). Therefore, it does not seem that the composition of the amino acid contributes to the codon and base biases in the boundary regions. Moreover, each of the three most frequent codons, AAA, AAU, and AGU, indicates the highest *G*-term in the second codon, while the order of appearance frequency does not match the *G*-term in other codon positions. These data suggest that the codon biases in the second codon have been generated in a distinct process from that of codon biases in other positions, which have been influenced by elongation steps, decoding by tRNA (*30, 31*). The statistical significance of codon biases was investigated by experimental studies on the effects of the sequence after the initiation codon on translation efficiency.

In the *in vitro* and *in vivo* assay systems used in this study, the statistically favorite sequences at the second codon exhibited positive effects on translation efficiency in *E. coli*. Our results agree with the previous report, in which it was proposed that the expression level of the *lacZ* gene is effected positively by the highly frequent second codon in several *E. coli* genes (*32*). When does the second codon influence translation efficiency? There are three major phases of translation, initiation, elongation, and termination. We do not discuss the termination steps here, even if they are worth noting. If the second codon affects the elongation step, from the initiation codon to the second codon or the second to the third codon, the optimal codon should be used at this position, as mentioned above. The favorite codon in the second position is not always the optimal codon. In this case, probably the initiation steps are enhanced by sequences in the second codon. We found the sequence AUGAAU, which possesses a notable efficiency, to be identical to a part of the downstream box (DB), which is usually located 9 to 12 nucleotides downstream of the initiation codon as the up regulator for translation. Although the mechanism has not yet been elucidated, recent studies

show that DB may play a role in the enhancement of gene expression together with the SD sequence, rather than alone (*23, 33*). Even if the internal AUG sequence in the DB serves as the initiation codon, the effect of translation enhancement appears strongly (*23*). Therefore, we propose that the sequence AUGAAU, deduced statistically, might be the DB and positively effect the formation of the translation initiation complex. However, the translation efficiency was decreased by having the third codon constituting the DB sequences. The results show that the effect of the strong complementarity in this region to anti-DB in the 16S rRNA appears to differ from that of usual DB. A stronger complementary DB causes more efficient translation (*20*). It is unclear that the sequence AUGAAU in this study and the so-called DB make the same contribution to translation. How do the sequences downstream of the initiation codon promote translational initiation? The relative efficiency differs among synonymous codons (Figs. 5 and 7), and the superiority of the translation efficiency is not changed by mutagenesis in the 16S rRNA (Fig. 9). These results provide evidence that base pairing between DB and anti-DB does not control translation. There has been disagreement as to whether DB interacts with anti-DB directly (*24, 34, 35*). One conclusion is that after the formation of the initiation complex, there is no direct interaction between DB and anti-DB (*33*). The ribosomal 30S subunit forms a pre-initiation complex before the association with the 50S subunit as follows. The SD sequence in mRNA first forms base pairs with the anti-SD in the 16S rRNA, then the anticodon of the initiator $tRNA^{Met}_f$ carried by initiation factor 2 interacts with the initiation codon. As demonstrated by toe-print experiments, the mRNA is fixed on the 30S subunit in the latter step (*36*). It is possible to anticipate that the mRNA itself will have an effect on translation when the interaction of the 30S particle and mRNA occurs loosely in the first step, or when the mRNA is rearranged in the 70S complex following association with the 50S subunit. The anti-DB region of the 16S rRNA has been proposed to be important for the 30S subunit to associate with the 50S subunit (*37*). Indeed, the greater the number of mutations in the anti-DB region (Fig. 8), the more slowly the transformants grow (data not shown). This suggests that the anti-DB region may be important for the formation of the initiation complex, resulting in a lower rate of cell growth.

Other possibilities concerning the effects on translation efficiency include the secondary structure of the mRNA around the initiation codon and mRNA binding proteins, for example, ribosomal protein S1. The amino terminal domain of S1 is associated with the 30S subunit by protein–protein interactions, whereas the elongated and flexible carboxy-terminal domain, containing RNA-binding centers with a high affinity for pyrimidine-rich sequences, provides the ribosome with an mRNA catching arm (*38*). The *E. coli* 30S ribosomal subunit recognizes the translational start sites of mRNAs lacking the SD sequence and forms stable ternary complexes only in the presence of S1 and IF3 (*4*). The previous report showed that ribosomal protein S1 recognizes a higher ordered structure of RNA (*39*). There is a possibility that the sequences downsteam of the initiation codon accelerates the interaction of mRNA with S1, so that the translation efficiency is enhanced by the formation of the initiation complex. Although there is no notable difference among the secondary structures of the DHFR mRNAs used here (data not shown), a slight structural difference in the region close to the initiation codon might lead to a proper interaction of mRNA and the ribosome. It remains uncertain which sequence or structure other than the SD sequence is advantageous for the ribosome to form an initiation complex in general. Table II shows the top five translationally efficient second codons in the *in vitro* assay. There is superiority when the first or second base of the second codon is a purine base, especially adenine, and the third base is adenine or uracil. These nucleotides would have been selected during the evolution of the translational machinery.

TABLE II. **Highest values of relative efficiency obtained in the *in vitro* assay of *dhfr* mutants in the second codon.**

Second codon	Translational activity
AUC (wild type)	1.00
AAU	2.95 ± 0.41
AAA	2.40 ± 0.93
AGU	2.23 ± 0.46
AUA	2.22 ± 0.43
GAU	1.88 ± 0.52

Recently, crystallographic structures of the bacterial 30S and 50S ribosomal subunits have been resolved at atomic resolution (*40–43*). Crystals of the complex of the 70S ribosome, mRNA and tRNA were also analyzed (*44*). These results are shedding light on the functional correlation of rRNAs and ribosomal proteins. Further understanding of ribosomal function based on structure will reveal the present translational apparatus, then the process of evolution can be clarified in the near future.

We are grateful to Dr. Michael O'Conner and Dr. Albert E. Dahlberg, J. W. Wilson Laboratory, Brown University, for providing pKK1192U, and Dr. Kimitsuna Watanabe, Graduate School of Frontier Sciences, The University of Tokyo, for providing pTHL5-12.

REFERENCES

1. Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**, 1–45
2. Shine, J. and Dalgarno, L. (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346
3. Steitz, J.A., and Jakes, K. (1975) How ribosomes select initiator regions in mRNA: Base pair formation between the 3′-teminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **72**, 4734–4738
4. Tzareva, N.V., Makhno, V.I., and Boni, I.V. (1994) Ribosome-messenger recognition in the absence of the Shine-Dalgarno interactions. *FEBS Lett.* **337**, 189–194
5. Philippe, C., Eyermann, F., Bénard, L., Portier, C., Ehresmann, B., and Ehresmann, C. (1993) Ribosomal protein S15 from *Escherichia coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc. Natl. Acad. Sci. USA* **90**, 4394–4398
6. Philippe, C., Bénard, L., Eyermann, F., Cachia, C., Kirillov, S.V., Portier, C., Ehresmann, B., and Ehresmann, C. (1994) Structural elements of *rpsO* mRNA involved in the modulation of translational initiation and regulation of *E. coli* ribosomal protein S15. *Nucleic Acids Res.* **22**, 2538–2546
7. van Duin, J. (1988) Single-strand RNA Bacteriophages in *The*

Bacteriophages (Calender, R., ed.), Vol. 1, pp. 117–167, Plenum Press, New York
8. Watanabe, H., Gojobori, T., and Miura, K. (1997) Bacterial features in the genome of *Methanococcus jannaschii* in terms of gene composition and biased base composition in ORFs and their surrounding regions. *Gene* **205**, 7–18
9. Watanabe, H., Gojobori, T., Terabe, M., Wakiyama, M., and Miura, K. (1997) Characteristic distribution of bases and codons around the initiation and termination codons in whole reading frames in bacteria and yeast genomes. *Nucleic Acids Symp. Ser.* **37**, 297–298
10. Poole, E.S., Brown, C.M., and Tate, W.P. (1995) The identity of the base following the stop codon determines the efficiency of *in vivo* translational termination in *Escherichia coli*. *EMBO J.* **14**, 151–158
11. Poole, E.S., Major, L.L., Mannering, S.A., and Tate, W.P. (1998) Translational termination in *Escherichia coli*: three bases following the stop codon crosslink to release factor 2 and affect the decoding efficiency of UGA-containing signals. *Nucleic Acids Res.* **26**, 954–960
12. Tohda, H., Chikazumi, N., Ueda, T., Nishikawa, K., and Watanabe, K. (1994) Efficient expression of *E. coli* dihydrofolate reductase gene by an in vitro translation system using phosphorothionate mRNA. *J. Biotechnol.* **34**, 61–69
13. Brosius, J., Ullrich, A., Raker, M.A., Gray, A., Dull, T.J., Gutell, R.R., and Noller, H.F. (1981) Construction and fine mapping of recombinant plasmids containing the *rrnB* ribosomal RNA operon of *E. coli*. *Plasmid* **6**, 112–118
14. Sigmund, C.D., Ettayebi, M., Borden, A., and Morgan E.A. (1988) Antibiotic resistance mutations in ribosomal RNA genes of *Escherichia coli*. *Methods Enzymol.* **164**, 673–690
15. Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**, 389–409
16. Grosjean, H. and Fiers, W. (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* **18**, 199–209
17. Harada, F. and Nishimura, S. (1974) Purification and characterization of AUA specific isoleucine transfer ribonucleic acid from *Escherichia coli* B. *Biochemistry* **13**, 300-307
18. Muramatsu, T., Yokoyama, S., Horie, N., Matsuda, A., Ueda, T., Yamaizumi, Z., Kuchino, Y., Nishimura, S., and Miyazawa, T. (1988) A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. *J. Biol. Chem.* **263**, 9261–9267
19. Sprengart, M.L., Fatscher, H.P., and Fuchs, E. (1990) The initiation of translation in *E. coli*: apparent base pairing between the 16S rRNA and downstream sequences of the mRNA. *Nucleic Acids Res.* **18**, 1719–1723
20. Sprengart, M.L., Fuchs, E., and Porter, A.G. (1996) The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. *EMBO J.* **15**, 665–674
21. Sprengart, M.L., and Porter, A.G. (1997) Functional importance of RNA interactions in selection of translation initiation codons. *Mol. Microbiol.* **24**, 19–28
22. Winzeler, E. and Shapiro, L. (1997) Translation of the leaderless *Caulobacter dnaX* mRNA. *J. Bacteriol.* **179**, 3981–3988
23. Etchegaray, J.P. and Inouye, M. (1999) Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. *J. Biol. Chem.* **274**, 10079–10085
24. O'Conner, M., Asai, T., Squires, C.L., and Dahlberg, A.E. (1999) Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. *Proc. Natl. Acad. Sci. USA* **96**, 8973–8978
25. Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**, 13–34
26. Kurland, C.G. (1991) Codon bias and gene expression. *FEBS Lett.* **285**, 165–169
27. Movva, N.R., Nakamura, K., and Inouye, M. (1980) Amino acid sequence of the signal peptide of ompA protein, a major outer membrane protein of *Escherichia coli*. *J. Biol. Chem.* **255**, 27–29
28. Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **146**, 1–21
29. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292
30. Osawa, S., Muto, A., Ohama, T., Andachi, Y., Tanaka, R., Yamao, F. (1990) Prokaryotic genetic code. *Experientia* **46**, 1097–1106
31. Kanaya, S., Yamada, Y., Kudo, Y., and Ikemura, T. (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238**, 143–155
32. Looman, A.C., Bodlaender, J., Comstock, L.J., Eaton, D., Jhurani, P., de Boer, H.A., and van Knippenberg, P.H. (1987) Influence of the codon following the AUG initiation codon on the expression of a modified *lacZ* gene in *Escherichia coli*. *EMBO J.* **6**, 2489–2492
33. Resch, A., Tedin, K., Grundling, A., Mündlein, A., and Bläsi, U. (1996) Downstream box-anti-downstream box interactions are dispensable for translation initiation of leaderless mRNAs. *EMBO J.* **15**, 4740–4748
34. Yu, D.C., Wang, A.L., Botka, C.W., and Wang, C.C. (1998) Protein synthesis in *Giardia lamblia* may involve interaction between a downstream box (DB) in mRNA and an anti-DB in the 16S-like ribosomal RNA. *Mol. Biochem. Parasitol.* **96**, 151–165
35. La Teana, A., Brandi, A., O'Connor, M., Freddi, S., and Pon, C.L. (2000) Translation during cold adaptation does not involve mRNA-rRNA base pairing through the downstream box. *RNA* **6**, 1393–1402
36. Hartz, D., McPheeters, D.S., Traut, R., and Gold, L. (1988) Extension inhibition analysis of translation initiation complexes. *Methods Enzymol.* **164**, 419–425
37. Mitchell, P., Osswald, M., and Brimacombe, R. (1992) Identification of intermolecular RNA cross-links at the subunit interface of the *Escherichia coli* ribosome. *Biochemistry* **31**, 3004–3011
38. Gribskov, M. (1992) Translational initiation factors IF-1 and eIF2-α share an RNA-binding motif with prokaryotic ribosomal protein S1 and polynucleotide phosphorylase. *Gene* **119**, 107–111
39. Ringquist, S., Joens, T., Snyder, E.E., Gibson, T., Bon, I., and Gold, L. (1995) High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. *Biochemistry* **34**, 3640–3648
40. Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920
41. Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930
42. Wimberly, B.T., Brodersen, D.E., Clemons, Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339
43. Brodersen, D.E., Clemons Jr, W.M., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T., and Ramakrishnan, V. (2000) The structural basis for the action of the antibiotics tetracydine, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **103**, 1143–1154
44. Cate, J.H., Yusupov, M.M., Yusupova, G.Z., Earnest, T.N., and Noller, H.F. (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science* **285**, 2095–2104