Gene Ther Mol Biol Vol 10, 1-12, 2006

Low-usage codons and rare codons of *Escherichia coli*

Mini Review

Dequan Chen* and Donald E. Texada

Department of Ophthalmology, Louisiana State University Health Sciences Center, Shreveport, LA 71130

***Correspondence**: Dequan Chen, Ph.D., Institute for Retina Research, 8210 Walnut Hill Lane, PBI, Suite 010, Dallas, TX 75231, USA; Phone: (214) 345-6801; email: dequan.chen@irrdallas.org.

Key words: low-usage codon, rare codon, common codon, major codon, codon usage, infrequent codon, rare tRNA, codon optimization, and rare tRNA supplementation

Abbreviations: *Escherichia coli, (E. coli)*; rare codon, (RC); rare codon cluster, (RCC)

Received: 4 September 2005; Accepted: 28 December 2005; electronically published: January 2006

Summary

In *Escherichia coli (E. coli)*, a low-usage codon is defined as a codon that is used rarely or infrequently in the genome with usage frequency lower than the smallest value (or frequency cut-off) among the usage frequencies of non-degenerate codons (Met codon AUG and Trp codon UGG) and the optimal codons for amino acids Leu, Ile, Val, Ser, Pro, Thr, Ala, Arg, Gly and Gln that have 2 or more degenerate codons with each having specific corresponding cognate tRNA for the optimal codon. A rare codon (RC), an infrequent codon or a minor codon is equivalently defined as a synonymous codon or a stop codon that is not only used rarely or infrequently in a genome but also decoded by a low-abundant tRNA (rare tRNA) or other factor(s) in an organism. The translational rate for a sense RC is much lower than that for a common (major) codon due to tRNA availability. A low-usage codon is not necessarily a RC, e.g., Cys codons UGU and UGC, Thr codons ACU and ACG, or His codons CAC and CAU are not rare codons of *E. coli*. However, a RC is definitely a low-usage codon. In *E. coli,* there are about 30 low-usage synonymous sense codons but only 20 of them are determined to be the bacterial RCs including 7 (AGG, AGA, CGA, CUA, AUA, CCC and CGG) used at a frequency of < 0.5% (Group I) and 13 (ACA, CCU, UCA, GGA, AGU, UCG, CCA, UCC, GGG, CUC, CUU, UCU and UUA) used at a frequency of > 0.5% (Group II). Studies have demonstrated that all the RCs in Group I and the first 6 RCs in Group 2 can cause translational problems in *E. coli*.

I. Introduction

Many proteins including those that can be used in treatment of certain disease (e.g., insulin), can rarely be obtained in large quantities from their natural sources. Besides, their purification or isolation is often not easy, and the cost is often pretty high. Recombinant DNA techniques have been successfully used in the past to express and purify these kinds of proteins. The bacterium *E. coli* has been and will continue to be the main, popular and first-choice expression host because it facilitates recombinant protein expression by its relative simplicity, its inexpensive and fast high-density cultivation, its well-known genetics and the availability of a large number of compatible tools including mutant strains, recombinant fusion partners and plasmids (Gold, 1990; Hodgson, 1993; Olins and Lee, 1993; Kane, 1995; Makrides, 1996; Jonasson et al, 2002; Sorensen and Mortensen, 2005a; Sorensen and Mortensen, 2005b). However, not every foreign gene can be efficiently expressed in *E. coli*, probably due to the unique and subtle structure of the target gene, the mRNA low stability and slow translational efficiency, the uneasy protein folding, the target protein degradation by *E. coli* proteases, the different codon usage between the organism of the foreign gene and native *E. coli,* or the toxicity of the expressed target protein (Olins et al, 1993; Makrides, 1996; Jonasson et al, 2002).

A number of studies have revealed that RCs and rare codon clusters (RCC) are capable of qualitatively and quantitatively causing expression problems in *E. coli* or other organisms (Kane, 1995; Makrides, 1996; Gurvich et al, 2005), and these problems mainly occur on translation level rather than on transcription level or other levels. The main translational problems caused by RCs or RCCs include (a) that rare codons reduce the translation rate of the target gene, (b) the expressed target protein is low or undetectable, (c) amino acids are misincorporated into the target protein, (d) truncated or amino acids-deleted peptides or proteins are synthesized, and (e) frame-shifted peptides or proteins are synthesized (Pedersen, 1984; Pohlner et al, 1986; Sorensen et al, 1989; Gurskii et al, 1992b; Kane et al, 1992; Gursky and Beabealashvilli, 1994; Vilbois et al, 1994; Kane, 1995; Calderone et al, 1996; Kleber-Janke and Becker, 2000; Kapust et al, 2002; McNulty et al, 2003; Flick et al, 2004; Shu et al, 2004; Choi et al, 2004; Chen et al, 2004; Gurvich et al, 2005). However, different groups often arbitrarily used different sets of codons as their rare or low-usage codons, or equivalently used low-usage codons and rare-tRNA associated codons. This may at least result in the following problems: (a) some codons are rare codons or low-usage codons to some groups but not to others, and vice versa; and (b) over- or under-estimation of the effects of rare codons on the expression of a gene even in the same system just because of the difference of low-usage or rare codons being defined or studied. To overcome these problems, universal meanings for low-usage codon and/or rare codon should be defined and the list of low-usage codons or rare codons in an organism should be determined. Therefore, the objectives of this review are mainly to unify and differentiate the meanings of a low-usage codon and a rare-tRNA associated codon (RC in short) as well as to determine the lists of the low-usage codons and rare-tRNA associated codons in *E. coli.*

II. Codon usage in *E. coli*

Codon usage was defined by Zhang et al in 1999 as the number of times (frequency) a codon is translated per unit time in the cell of an organism. This is a definition for real-time codon usage. But it is hard to be measured in vivo. Zhang et al, used 3 different methods to estimate the codon usage in *E. coli* and other organisms in their studies including measuring the average frequencies of codons in the sequenced protein-coding genes in an organism. All their methods gave approximately the same results as regards the hierarchy for "most used' and "least used" codons within each synonymous codon family (Zhang et al, 1991). Therefore, it is reasonable to use averaged codon frequency of the sequenced protein-coding reading frames of an organism to roughly represent the real-time codon usage although this may over-estimate the usages of infrequently or rarely used codons and underestimate those of frequently used codons because different reading frames are used and translated for different number of times in the organism at a given time (Zhang et al, 1991). Besides, this is what codon usage generally means to many scientists in the past and at present.

Before the 1980s and after the discovery of genetic code redundancy or degeneracy (an amino acid except Met and Trp is encoded by 2 to 6 codons), it was often thought that degenerate codons for the same amino acid were used randomly in a genome. This is based on the simplest assumption that all genomes have uniform codon usage meaning that synonymous codons (degenerate codons for same amino acid) are used with equal frequency. As more and more sequence data (especially the gene sequences of bacterium *E. coli and* yeast *Saccharomyces cerevisiae*) appeared in the late 1970s and early 1980s, it came to light that (a) synonymous codon usage is consistently similar for all genes within each type of genome or organism (Grantham et al, 1980a,b, 1981), and (b) synonymous codon usage was not random, i.e., synonymous codons are not used with equal frequency in a genome (Ikemura, 1985; Sharp et al, 1988; Zhang et al, 1991; Sorensen et al, 2005a). This is also true for non-synonymous codon usage (non-random usage of different codons for different amino acids). Therefore, the codon usage among degenerate codons in each organism is biased, with some codons more preferred (at higher usage frequency or used more frequently) than the other(s). Further, studies also found that codon usage bias is greater in highly expressed genes than poorly expressed genes (Gouy and Gautier, 1982; Sharp and Li, 1986; Makrides, 1996). That is to say, highly expressed genes in an organism mostly use preferred codons (especially the most preferred or optimal codons) and avoid non-preferred codons while poorly expressed genes use fewer preferred or optimal codons but more non-preferred codons (Ikemura, 1985). Meanwhile, codon pair usage was even found not to be random (Nussinov, 1981; Lipman and Wilbur, 1983; Gutman and Hatfield, 1989; Irwin et al, 1995).

The codon usage frequencies for the 64 codons (3 stop codons, and 61 sense codons - codons that encode amino acids) of *E. coli*, calculated from the GenBank genetic sequence data (Releases # 63, 69 and 147), are shown in **Table 1**. The data in the table demonstrate that as the total number of the codons or protein-coding genes included in each GenBank release increases (especially from #69 to #147, which is about 8 times increase), the calculated frequency for a given sense codon changes as follows:

(a) The frequencies of low-usage codons (highlighted by red, purpurple and blue) has a tendency to increase (those of CUC, GUA, UCG, CCA, CAU, UGU, CGA, UCU and ACU increase very little while those of others a lot) except those of CAC, UGC and UCC (the last two usage frequencies decrease very little);

(b) The frequencies of some high-usage codons change very little (those of GUA, ACU, GCC, GCA and GAU increase while those of AUG, GUU, GUC, GCU, AAA, GAG and AGC decrease),

(c) The frequencies of some high-usage codons have a tendency to increase (for those of UUU, AUU, UAU, CAA, AAU and UGG) while the frequencies of other high-usage codons, on the contrary, have a tendency to decrease (for those of UUC, CUG, AUC, GUG, CCG, ACC, GCG, UAC, CAG, AAC, GAC, GAA, CGU, CGC GGU and GGC).

The above results may imply the following:

(1) Some codons, whether at high-usage (see above b) or at low-usage (see above a), are used at about the same frequency in the old sequenced proteins (e.g., the proteins included in GenBank release #69) as in the new sequenced proteins (e.g., the proteins included in GenBank release #147 but not in #69). Therefore, their usage-frequencies change very little between the GenBank releases, and the usage frequency calculated from GenBank release #69 or 147 should all well represent their real-time codon usage frequencies (**Table 1**).

(2) Most low-usage codons (see above a) and some high-usage codons (see above c) are not well used by the old sequenced proteins, and the new sequenced proteins (e.g., the proteins included in GenBank release #147 but

not in #69) have used more of them. Therefore, their usage frequencies increase over the total number of protein genes included in the GenBank releases. Two factors should contribute to the phenomenon: one factor is that these codons especially low-usage codons are more frequently used in the new sequenced protein genes (most of them are poorly expressed), which results in the increase of their calculated usage frequencies; the other factor, on the contrary, is that poorly expressed genes have low expression rates at a give time of the bacterial life, and averaging over the entire genome without weighting the number of times different reading frames are being translated leads to over-estimation of their codon usage frequencies (Zhang et al, 1991). Therefore, the real-time codon usage frequencies for these codons should be much lower than the data calculated from GenBank release #147 but somewhere around the data from GenBank release #69 (**Table 1**).

(3) Some high-usage codons (see above c) are well used by the old sequenced proteins, and the new sequenced proteins (e.g., the proteins included in GenBank release #147 but not in #69) have used less of them. Therefore, their usage frequencies decrease over the total number of protein genes included in the GenBank releases. The above two factors should also contribute to this but in a reverse direction: the first factor is that these codons are less frequently used in the new sequenced protein genes (most of them are poorly expressed), which results in the true decrease of their calculated usage frequencies; the second is that poorly expressed genes have low expression rates at a give time of the bacterial life, and averaging over the entire genome without weighting the number of times different reading frames are being translated leads to under-estimation of the codon usage frequencies for these codons (Zhang et al, 1991). Therefore, the real-time codon usage frequencies for these codons should be much higher than the data from GenBank release #147 but somewhere around the data from GenBank release #69 or even #63 (**Table 1**).

The above analysis suggests that the frequency values listed in the II columns (calculated from GenBank release #69) of **Table 1** most likely and approximately better represent the real-time codon usage frequencies in *E. coli.*

Dong et al, measured *E. coli* codon usage frequencies at different bacterial growth rates (0.4-2.5 doublings per hour), which were calculated from the coding frames of 140 protein mRNAs (Dong et al, 1996). The results has been adapted and presented to **Table 2**.

Table 1. Codon frequencies used by protein-coding reading frames of *E. coli*[a]

	I [b]	II [c]	III [d]		I [b]	II [c]	III [d]		I [b]	II [c]	III [d]		I [b]	II [c]	III [d]
UUU	18.85	19.2	22.46	UCU	10.47	10.4	10.94	UAU	15.09	15.4	18.34	UGU	4.80	4.7	5.35
UUC	18.07	18.2	15.62	UCC	9.43	9.4	9.29	UAC	13.29	13.4	12.01	UGC	6.07	6.1	5.99
UUA	10.52	10.9	14.98	UCA	6.52	6.8	9.94	UAA	1.99	2.0	1.99	UGA	0.80	0.8	1.04
UUG	11.33	11.5	12.86	UCG	7.89	8.0	8.52	UAG	0.20	0.2	0.29	UGG	12.90	12.8	13.78
CUU	9.92	10.2	12.49	CCU	6.57	6.6	7.90	CAU	11.35	11.6	12.47	CGU	24.70	24.1	18.92
CUC	9.70	9.9	10.08	CCC	4.19	4.3	5.63	CAC	10.74	10.7	8.82	CGC	21.50	22.1	18.38
CUA	2.97	3.2	4.47	CCA	8.12	8.2	8.63	CAA	13.07	13.2	14.38	CGA	3.06	3.1	4.03
CUG	54.10	54.6	46.04	CCG	23.91	23.8	19.35	CAG	29.68	30.1	28.12	CGG	4.62	4.6	6.49
AUU	27.27	27.2	29.67	ACU	10.83	10.2	11.02	AAU	16.30	16.3	22.83	AGU	7.37	7.2	10.73
AUC	26.97	26.5	22.69	ACC	24.37	24.3	21.39	AAC	24.35	23.9	21.20	AGC	14.95	15.2	15.00
AUA	3.94	4.1	8.22	ACA	6.53	6.5	10.70	AAA	37.47	36.5	35.60	AGA	2.14	2.1	4.47
AUG	26.33	26.5	25.95	ACG	12.54	12.7	13.78	AAG	11.94	12.0	13.05	AGG	1.32	1.4	2.56
GUU	20.79	20.1	20.04	GCU	17.86	17.4	17.36	GAU	32.14	32.3	32.88	GGU	28.48	27.6	24.93
GUC	14.09	14.2	14.04	GCC	23.18	23.5	23.87	GAC	22.03	21.8	18.83	GGC	30.41	30.2	25.66
GUA	12.06	11.6	11.90	GCA	20.92	20.8	21.60	GAA	43.75	43.4	38.02	GGA	6.95	7.0	10.61
GUG	24.68	25.3	23.47	GCG	32.94	33.1	27.99	GAG	19.03	19.2	18.80	GGG	9.63	9.7	11.58

a. The usage of each codon is expressed as the frequency per 1000 codons, which is calculated by division of the absolute number of the indicated codon by the total number of codons used in all the sequenced *E. coli* protein-coding sequences or reading frames.

b. Taken from Zhang et al (1991). Codon usage frequency was calculated from 323059 codons of 968 protein coding reading frames (CDS) (GenBank Version 63.0, 15 March 1990).

c. Taken from Wada et al (1991). Codon usage frequency was calculated from 524410 codons of 1562 protein coding reading frames (CDS) (GenBank Version 69.0, September 1991).

d. Taken and adapted from http://www.kazusa.or.jp/codon (Nakamura et al, 2000). Codon usage frequency was calculated from 4182749 codons of 13778 protein coding reading frames (CDS) (GenBank Version 147.0, 1 June 2005).

Table 2. Real-time codon frequencies used by protein-coding reading frames of *E. coli*[a]

	Growth Rate [b]				Growth Rate [b]				Growth Rate [b]				Growth Rate [b]		
	0.4	1.07	2.5		0.4	1.07	2.5		0.4	1.07	2.5		0.4	1.07	2.5
UUU	12.55	10.30	7.92	UCU	13.12	14.14	16.33	UAU	10.68	8.90	6.72	UGU	4.23	3.64	2.76
UUC	22.68	22.44	23.25	UCC	11.15	12.09	11.68	UAC	16.20	16.71	16.52	UGC	5.29	4.77	3.81
UUA	6.13	4.64	2.73	UCA	3.89	3.09	1.98	UAA	2.77	3.38	4.18	UGA	0.31	0.23	0.19
UUG	6.63	5.72	4.27	UCG	6.05	4.58	2.51	UAG	0.00	0.00	0.00	UGG	9.76	8.69	7.03
CUU	5.70	4.64	3.86	CCU	4.99	4.79	4.38	CAU	9.23	8.11	6.78	CGU	31.12	36.61	43.82
CUC	6.19	5.52	4.09	CCC	3.32	2.10	1.09	CAC	13.90	13.91	14.21	CGC	22.25	22.39	20.59
CUA	2.15	1.53	0.82	CCA	6.52	6.40	5.18	CAA	10.91	8.98	7.01	CGA	1.32	0.99	0.67
CUG	60.13	61.29	60.75	CCG	29.51	28.88	28.82	CAG	29.24	28.33	27.28	CGG	1.75	1.23	0.62
AUU	21.38	19.26	15.79	ACU	13.88	16.76	20.64	AAU	9.79	7.79	5.61	AGU	3.99	3.01	2.19
AUC	36.68	39.15	43.86	ACC	26.51	27.10	26.70	AAC	27.95	28.64	29.21	AGC	11.97	10.69	9.31
AUA	0.93	0.75	0.52	ACA	3.48	2.99	2.61	AAA	44.43	49.07	55.01	AGA	1.12	0.84	0.63
AUG[c]	25.32	25.82	25.90	ACG	7.53	6.21	4.17	AAG	12.08	13.74	17.22	AGG	0.09	0.05	0.03
GUU	31.31	35.63	43.18	GCU	28.85	32.14	39.49	GAU	24.25	22.40	19.27	GGU	38.29	40.49	45.55
GUC	11.25	9.71	7.67	GCC	19.80	16.81	11.81	GAC	28.72	30.93	33.74	GGC	35.62	35.54	34.17
GUA	15.87	18.65	22.31	GCA	22.13	22.38	24.87	GAA	53.10	55.10	57.86	GGA	2.71	2.21	1.26
GUG	21.40	18.93	14.98	GCG	30.33	28.45	24.11	GAG	16.57	17.04	16.97	GGG	4.81	3.57	2.36

a. Taken from Dong et al (1996). The usage of each codon is expressed as the frequency per 1000 codons. The codon frequencies were the averages from 140 proteins and calculated on the basis of the relative weight fraction of each protein and on the assumption that the amount of a protein accumulated in the cell during the steady growth is proportional to the amount of its corresponding mRNA in the bacteria.
b. Growth rate is expressed as doublings per hour. Different growth rates were obtained by varying the nutrient contents of the culturing media (Dong et al, 1996).
c. The data for AUG usage frequency are the sum of the frequencies for Met_{f1}, Met_{f2}, and Met_{m}.

Although the number of protein coding frames used is very small, the frequency values were obtained by weighting every protein amount at each growth rate of *E. coli* according to the data reported by Pedersen et al (Pedersen et al, 1978). Therefore, the codon usage frequencies in **Table 2** are real-time codon usage values. The data in **Table 2** demonstrate that (a) *E. coli* codon usage is biased at all studied bacterial growth rate, (b) the frequencies of low-usage sense codons (marked by red and purple) decrease with increasing growth rate, and (c) the frequencies of some high-usage sense codons (UUC, AUC, GUU, GUA, UCU, ACU, GCU, GCA, CAC, AAC, AAA, GAC, GAA, CGU and GGU) increase while those of others decrease over the increase of growth rate. In addition, most codon usage frequencies in **Table 2** are in good agreement with those in **Table 1**.

III. tRNA abundance in *E. coli*

Codon usage bias in an organism may have been formed during evolution by the combinatory effects of various factors such as the adaptation of gene expressivity to various growth conditions (Gouy et al, 1982), the adaptation of codons to tRNA availability (Ikemura, 1980, 1981a,b, 1985), the adaptation of codon-anticodon paring or interaction to have optimal or intermediate energy strength (Grosjean et al, 1978; Grosjean and Fiers, 1982), the adaptation of codon mutations to form specific mRNA secondary structure(s), etc. But codon adaptation to tRNA availability are attributed to have played a key role in the formation of biased codon usage because organism-specific codon usage patterns were demonstrated to correlate with the abundance spectra of organism-specific populations of isoaccepting or cognate tRNAs (Ikemura, 1980, 1981a,b, 1985).

The relative contents of tRNAs for normally growing *E. coli*, which were measured by Ikemura (1981a, 1981b, 1985), are listed in **Table 3**. The data (relative contents for 38 or 40 tRNAs) in the table demonstrate that: (a) the abundance of $\text{tRNA}^{\text{Gly3}}$ is the highest (relative amount is 1.1) among all the *E. coli* tRNAs and it can recognize/decode two codons (GGU and GGC), immediately followed by $\text{tRNA}^{\text{Val1}}$, $\text{tRNA}^{\text{Ala(GCY)}}$ and

tRNA^{Ile1} (relative amounts are 1.05, 1.04 and 1.0 respectively) in succession with the first recognize 3 codons (GUA, GUG and GUU) while the latter 2 recognizing 2 codons (GCC and GCU, AUU and AUC respectively); (b) tRNA^{Leu1} is a tRNA that recognizes only one single codon (CUG) and at the same time has the highest abundance (relative amount is 1.0); (c) some tRNAs including the cognate tRNAs for CUA, AUA, CGG, AGA and AGG, ACA and ACG, CCC, or UGU and UGC, have very low abundances while the abundances for other tRNAs are different with relative amounts ranging from 0.1 to 0.9; and (d) UCU (for Ser), GUU (for Val), GCU (for Ala), and GGG (for Gly) are recognized by 2 isoacceptor-tRNAs. In addition, the relative contents of tRNAs (43 or 45 tRNAs) for *E. coli* growing at different rates (0.4, 0.7, 1.07, 1.6 and 2.5 doublings/hour), measured by Dong et al (Dong et al, 1996), are listed in **Table 4**.The data in **Table 4** suggest that tRNA abundance in *E. coli* varies with bacterial growth rate - increases over the increase of growth rate (the increase amplitude varies with different tRNAs). Most tRNA relative contents in **Table 4** are in agreement with those in **Table 3**. The data of **Tables 1** and 2, with those of **Tables 3** and **4**, altogether support the concept that the usage frequency of synonymous codon often reflects or correlates with the abundance of its cognate tRNA in *E. coli* (Garel, 1974; Garel et al, 1981; Ikemura, 1985; Bulmer, 1987; Emilsson and Kurland, 1990; Emilsson et al, 1993; Kane, 1995; Makrides, 1996; Dong et al, 1996).

IV. Definition of low-usage codon and rare codon

Low-usage codon is often called RC, minor codon, and infrequent codon because all of them imply that the usage of such a codon in a genome or an organism is low or very low, in other words, the codon is used rarely or infrequently in a genome or an organism. All the above terms have been equivalently used in the past. But different groups defined different sets of codons as their low-usage codons (although most groups included the several least usage codons in their low-usage codon sets) due to (a) the different numbers of the available protein-coding gene sequences for calculating the codon usage frequencies, and (b) the arbitrary frequency cut-offs which were used by different people (e.g. 0.5%, 1.0%, or 1.1%) to define the boundary between low-usage codons and common codons. The above may result in the following problems: (a) some codons, are low-usage codons to some people but not to others, and vice versa; e.g., GUC and GCC were considered as RC by Pedersen (1984) but not us; (b) different results or conclusions regarding the effects of low-usage codons on the expression of a gene(s) (often over- or under-estimation occurs) may be obtained for the same system just because of the difference of low-usage codons being defined or studied; (c) the results from different groups, for the same gene, are often hard to be compared with each other. Therefore, universal definition(s) for the above terms, or universal terms with fixed meanings is required.

The correlation of the usage frequency of a synonymous codon with its cognate tRNA abundance (such as high-usage codons with high-abundant tRNAs and low-usage codons with low-abundant tRNAs), together with the so far reported expression problems derived from low-usage codons and/or their cognate tRNA availability, suggest that just one term to cover all the above meanings is not enough. To satisfy the above requirements, a RC, an infrequent codon or a minor codon is equivalently defined as a synonymous codon that is not

Table 3. Relative contents of tRNAs in *E. coli* [a]

tRNA		Recognized codon	Content [b]
Leu:	1	CUG	1.00
	2	CUU, CUC	0.30
	UUR	UUA, UUG	0.25
	CUA	CUA	minor
Val:	1	GUA, GUG, GUU*	1.05
	2	GUC, GUU*	0.40
Gly:	1	GGG*	0.10
	2	GGA,GGG*	0.15
	3	GGU, GGC	1.10
Ala:	1	GCA, GCG, GCU*	0.85
	GCY	GCC, GCU*	1.04
Arg:	1, 2	CGU, CGC, CGA	0.90
	CGG	CGG	minor
	AGR	AGA, AGG	minor
Ile:	1	AUU, AUC	1.00
	2	AUA	0.05
Lys		AAA, AAG	1.00
Glu	2 (1)	GAA, GAG	0.90
Asp	1	GAU, GAC	0.80
Thr:	1+3	ACU, ACC	0.80
	4	ACA, ACG	minor
Asn		AAU, AAC	0.60
Gln:	1	CAA	0.30
	2	CAG	0.40
Tyr:	1+2	UAU, UAC	0.50
Ser:	1	UCU*, UCA, UCG	0.25
	3	AGU, AGC	0.25
	UCY	UCC, UCU*	
His		CAC, CAU	0.40
Trp		UGG	0.30
Pro:	1	CCG	major
	2	CCC	minor
	3	CCU, CCA, CCG	major
Phe		UUU, UUC	0.35
Cys		UGU, UGC	minor
Met:	m	AUG	0.30
	f1	AUG	0.40
	f2	AUG	0.10

a. Taken and adapted from Ikemura (1981a,b, 1985)
b. The content is the relative amount to that of $\mathrm{tRNA}^{Leu1(CUG)}$ that is normalized to 1.0 and approximately on the order of 10^4 molecules per cell for normally growing *E. coli.*
*. A single codon is recognized by 2 tRNAs.

Table 4. Relative contents of tRNAs in *E. coli* at different growth rates

tRNA		Recognized codon(s)	Growth Rate (doublings per hour)				
			0.4	0.7	1.07	1.6	2.5
Leu:	1	CUG	1.00	1.06	1.19	1.51	1.57
	2	CUC, CUU	0.21	0.25	0.29	0.33	0.42
	3	CUA, CUG	0.15	0.18	0.19	0.23	0.22
	4	UUG	0.43	0.45	0.49	0.68	0.66
	5	UUA,UUG	0.25	0.25	0.29	0.26	0.27
Val:	1	GUA, GUG, GUU	0.86	0.86	0.78	1.35	1.45
	2A	GUC, GUU	0.14	0.14	0.17	0.19	0.20
	2B	GUC, GUU	0.14	0.17	0.19	0.26	0.31
Gly:	1+2	(GGG) / (GGA,GGG)	0.48	0.51	0.55	0.78	0.79
	3	GGC, GGU	0.98	1.08	1.19	1.41	1.77
Ala:	1B	GCU, GCA, GCG	0.73	0.83	1.00	1.24	1.49
	2	GCC	0.14	0.15	0.17	0.23	0.25
Arg:	2	CGU, CGC, CGA	1.06	1.03	1.10	1.68	1.81
	3	CGG	0.14	0.18	0.10	0.16	0.16
	4	AGA	0.19	0.17	0.19	0.23	0.25
	5	AGG	0.09	0.11	0.11	0.17	0.16
Ile:	1+2	(AUC, AUU) / AUA	0.78	0.84	0.94	1.34	1.75
Lys		AAA, AAG	0.43	0.48	0.52	0.62	0.74
Glu	2	GAA, GAG	1.05	1.10	1.18	1.71	2.08
Asp	1	GAC, GAU ,	0.54	0.58	0.60	0.85	1.10
Thr:	1	ACC, ACU	0.02	0.03	0.04	0.04	0.05
	2	ACG	0.12	0.14	0.15	0.19	0.22
	3	ACC, ACU	0.25	0.26	0.27	0.34	0.39
	4	ACA, ACU, ACG	0.20	0.22	0.23	0.35	0.49
Asn		AAC, AAU	0.27	0.27	0.31	0.43	0.52
Gln:	1	CAA	0.17	0.19	0.26	0.22	0.31
	2	CAG	0.20	0.22	0.25	0.36	0.44
Tyr:	1	UAC, UAU	0.17	0.17	0.19	0.33	0.30
	2	UAC, UAU	0.28	0.27	0.27	0.37	0.36
Ser:	1	UCA, UCU, UCG	0.29	0.39	0.39	0.49	0.52
	2	UCG	0.08	0.07	0.08	0.10	0.10
	3	AGC, AGU	0.31	0.31	0.32	0.38	0.40
	5	UCC, UCU	0.17	0.18	0.20	0.26	0.29
His		CAC, CAU	0.14	0.16	0.19	0.24	0.31
Trp		UGG	0.21	0.20	0.24	0.29	0.36
Pro:	1	CCG	0.20	0.17	0.25	0.19	0.19
	2	CCC, CCU	0.16	0.18	0.16	0.28	0.27
	3	CCA, CCU, CCG	0.13	0.13	0.16	0.18	0.18
Phe		UUC, UUU	0.23	0.26	0.30	0.33	0.36
Cys		UGC, UGU ,	0.36	0.35	0.37	0.50	0.50
Met:	m	AUG	0.16	0.18	0.21	0.29	0.31
	f1	AUG	0.27	0.34	0.43	0.45	0.72
	f2	AUG	0.16	0.16	0.17	0.24	0.27

only used rarely or infrequently in a genome but also decoded by a low-abundant tRNA (rare tRNA) or other factor(s) such as less-efficient translation releasing factor(s) in an organism. Therefore, a RC encoding an amino acid is a rare-tRNA associated codon while a RC for translation termination is a stop codon with lowest usage frequency in a genome. Meanwhile, a low-usage codon is defined as a codon (whether synonymous or not) that is used rarely or infrequently in a genome, and its usage frequency should be: (a) lower than the usage frequencies of the non-degenerate codons (that is, AUG for Met, and UGG for Trp); (b) lower than the usage frequencies of the optimal codons for amino acids (Leu, Ile, Val, Ser, Pro, Thr, Ala, Arg, Gly and Gln) with 2 or more degenerate codons because these amino acids have 2 or more tRNA carriers with at least one to specify the corresponding optimal codon of each amino acid; (c) lower than the smallest value (cut-off frequency) among the usage frequencies listed in (a) and (b). Therefore, cut-off frequency is an objective value rather than an arbitrary one for defining the boundary between low-usage codons and common codons in each organism. Data in **Tables 1** and **2** suggest that the usage frequency of Trp codon UGG is the very cut-off frequency value of *E. coli.*

V. Determination of *E. coli* low-usage codons and rare codons

Based on the above definitions, a low-usage codon is not necessarily a RC but a RC is definitely a low-usage codon. All the 3 stop codons (UAA, UAG and UGA) of *E. coli* (**Tables 1 and 2**) are the least usage codons of the bacterium. According to the above definitions, they should be low-usage codons. However, UAA cannot be regarded as the rare stop codon but a major stop codon of *E. coli* because it has the highest usage among the 3 stop codons.

The low-usage sense codons of *E. coli*, based on the 1.28% cut-off of usage frequency calculated from GenBank release #69 (**Table 1**, columns of “II”) and on the 0.869% cut-off of real-time usage frequency calculated from 140 proteins of *E. coli* growing at the rate of 1.07 doublings/hour (**Table 2**), are all listed in **Table 5**. The relative tRNA contents measured by Ikemura (1980, 1981a,b, 1982, 1985) and Dong et al, 1996 are also included in the table. The 3 stop codons are all *E. coli* low-usage codons, but not listed in **Table 5**. Based on the above RC definition, 10 low-usage sense codons out of the 30 listed in **Table 5** are excluded from the list of *E. coli* rare sense codons because of the following reasons:

(a) UGU and UGC. They are the only synonymous codons of Cys, and both are decoded by a single tRNA^{Cys} which has a relative amount of > 0.36 (Dong et al, 1996).

(b) ACU and ACG. They are 2 synonymous codons of Thr (the total is 4), but they recognized by more than 2 tRNAs. ACU is recognized by tRNA^{Thr1}, tRNA^{Thr3} and tRNA^{Thr4}, and the sum of the relative contents of these 3 tRNAs is > 0.45 (Dong et al, 1996) or 0.8 (Ikemura, 1985). ACG is recognized by tRNA^{Thr2} and tRNA^{Thr4}, and the sum of the relative contents of these 2 tRNAs is > 0.32 (Dong et al, 1996) or minor (Ikemura, 1985).

(c) CAC and CAU. They are the only synonymous codons of His, and both are decoded by a single tRNA^{His} which has a relative amount of 0.4 (Ikemura, 1985) or > 0.14 (Dong et al, 1996).

(d) UUG. UUG is one of 6 synonymous codons of Leu, and is decoded by 2 tRNAs-tRNA$^{Leu4(UUG)}$ and tRNA$^{Leu5(UUA,UUG)}$. The relative amount of tRNA$^{Leu4(UUG)}$ is 0.43 (Dong et al, 1996) while tRNA$^{Leu5(UUA,UUG)}$ has a the relative amount of 0.25 (Ikemura, 1985) or > 0.25 (Dong et al, 1996).

(e) GUA. GUA is one out of the 4 synonymous codons of Val. The single tRNA that can decode this codon is very high abundant and the relative amount is 1.05 (Ikemura, 1985) or > 0.78 (Dong et al, 1996).

(f) AAG. Lys has only 2 synonymous codons AAA and AAG, and the 2 codons are decoded by a single tRNA^{Lys} which has a relative amount of 1.0 (Ikemura, 1985) or > 0.43 (Dong et al, 1996).

(g) AAU. Asn has only 2 synonymous codons AAU and AAC, and the 2 codons are decoded by a single tRNA^{Asn} which has a relative amount of 0.6 (Ikemura, 1985) or > 0.27 (Dong et al,1996). Besides, the usage frequencies of AAU calculated from the 3 GenBank releases are much higher than those (14.95, 15.2 and 15.0 per thousand) of AGC (the optimal codon of Ser).

Moreover, UGU and UGC for Cys, CAC and CAU for His, AAG for Lys, and AAU for Asn are not regarded as rare codon in **Table 5** because these amino acids all have 2 synonymous codons that are respectively decoded by a single tRNA (Crick’s “wobble hypothesis” can explain why a single tRNA can recognize multiple degenerate codons (Crick, 1966). In addition, whether the above 10 low-usage sense codons that have been excluded from *E. coli* rare codon list can cause significant expression problems, has never been reported.

There are 9 amino acids (Phe, Tyr, His, Gln, Asn, Lys, Asp, Glu and Cys) that have only 2 synonymous codons. **Tables 3** and **4** demonstrate that (a) the 2 synonymous codons of Phe, His, Asn, Lys, Cys, Asp or Glu are specified by a single tRNA; (b) the 2 codons (UAU and UAC) of Tyr are non-differentially recognized by the 2 tyrosinyl tRNAs (tRNA^{Tyr1} and tRNA^{Tyr2}); and (c) Gln has 2 tRNAs and each recognize a Gln codon (tRNA^{Gln1} for CAA and tRNA^{Gln2} for CAG). Although there exists usage difference in the 2 synonymous codons for each of the 8 amino acids Phe, Tyr, His, Asn, Lys, Asp, Glu and Cys, it is not due to their tRNA availability (Grosjean et al, 1978) but mainly to gene expressivity (Gouy et al, 1982; Ikemura, 1985). The usage differences in the 2 synonymous codons of the above 8 amino acids may be explained by the “rules” proposed for the choice or usage preference of the synonymous codons that are decoded by a single tRNA (Gouy et al, 1982; Ikemura, 1985). Therefore, any codon for the above amino acids (except Gln) cannot be regarded as a rare codon even though it may be a low-usage codon according to the usage cut-off determined as described above.

In this paper, the following codons are considered to be the rare sense codons of *E. coli* and are classified into 2 groups:

(a) Group I: AGG, AGA, CGA, CUA, AUA, CCC and CGG (arranged from least usage to high usage based on the average usage frequency). The usage frequency of each codon calculated from GenBank release #69 is < 0.5%.

(b) Group II: ACA, CCU, UCA, GGA, AGU, UCG, CCA, UCC, GGG, CUC, CUU, UCU and UUA (arranged from least usage to high usage based on the average usage frequency). The usage frequency of each codon calculated from GenBank release #69 is > 0.5% but < 1.1%.

Some rare sense codons in Group II (namely the first highlighted 6 in the above list) have been reported to be involved in translational problems (Konigsberg and Godson, 1983; Chen and Inouye, 1990a; Ma et al, 2003; Zhou et al, 2004). According to this, Group II can further be classified into 2 subgroups:

Group II$_a$: ACA, ACU, UCA, GGA, AGU and UCG;

Group II$_b$: CCA, UCC, GGG, CUC, CUU, UCU and UUA.

At present, whether rare sense codons in Group II$_b$ can cause expression problems have not been reported, and

Table 5. Low-usage sense codons of *E. coli*[a]

		Codon frequency (per thousand)[b]						Cognate tRNA relative amount and other codons[c]					
Amino		GenBank release #			Growth Rate			Amount[d]	Other Codons[e]	Growth Rate[f]			Other Codons[e]
Acid	Codon	63	69	147	0.4	1.07	2.5			0.40	1.07	2.50	
Arg	AGG	1.3	1.4	2.6	0.09	0.05	0.03	minor	AGA	0.09	0.11	0.16	/
Arg	AGA	2.1	2.1	4.5	1.32	0.99	0.67	minor	AGG	0.19	0.19	0.25	/
Arg	CGA	3.1	3.1	4	1.32	0.99	0.67	minor	CGU,CGC	1.06	1.1	1.81	CGU,CGC
Leu	CUA	3	3.2	4.5	2.15	1.53	0.82	minor	/	0.15	0.19	0.22	CUG
Ile	AUA	3.9	4.1	8.2	0.93	0.75	0.52	minor	/	0.78	0.94	1.75	AUC,AUU[g]
Pro	CCC	4.2	4.3	5.6	3.32	2.1	1.09	minor	/	0.16	0.16	0.27	CCU
Arg	CGG	4.6	4.6	6.4	1.75	1.23	0.62	minor	/	0.14	0.1	0.16	/
*Cys	UGU	4.8	4.7	5.3	4.23	3.64	2.76	minor	UGC	0.36	0.37	0.5	UGC
*Cys	UGC	6.1	6.1	6	5.29	4.77	3.81	minor	UGU	0.36	0.37	0.5	UGU
Thr	ACA	6.5	6.5	10.7	3.48	2.99	2.61	minor	ACG	0.2	0.23	0.49	ACG,ACU
Pro	CCU	6.6	6.6	7.9	4.99	4.79	4.38	major	CCG,CCA	0.16	0.16	0.27	CCC
	CCU									0.13	0.16	0.18	CCA,CCG
Ser	UCA	6.6	6.8	9.9	3.89	3.09	1.98	0.25	UCU,UCG	0.29	0.39	0.52	UCU,UCG
Gly	GGA	7	7	10.6	2.71	2.21	1.26	0.15	GGG	0.48	0.55	0.79	GGG[g]
Ser	AGU	7.4	7.2	10.7	3.99	3.01	2.19	0.25	AGC	0.31	0.32	0.4	AGC
Ser	UCG	7.9	8	8.5	6.05	4.58	2.51	minor	UCU,UCA	0.29	0.39	0.52	UCU,UCA
	UCG									0.08	0.08	0.1	/
Pro	CCA	8.1	8.2	8.6	6.52	6.4	5.18	major	CCG,CCU	0.13	0.16	0.18	CCG,CCU
Ser	UCC	9.4	9.4	9.3	11.2	12.1	11.7	/	UCU	0.17	0.2	0.29	UCU
Gly	GGG	9.6	9.7	11.6	4.81	3.57	2.36	0.15	GGA	0.48	0.55	0.79	GGA[g]
	GGG							0.1	/				/
Leu	CUC	9.7	9.9	10.1	6.19	5.52	4.09	0.3	CUU	0.21	0.29	0.42	CUU
Leu	CUU	9.9	10.2	12.5	5.7	4.64	3.86	0.3	CUC	0.21	0.29	0.42	CUC
§Thr	ACU	10.8	10.2	11	13.9	16.8	20.6	0.8	ACC	0.02	0.04	0.05	ACC
	ACU									0.25	0.27	0.39	ACC
	ACU									0.2	0.23	0.49	ACA,ACG
Ser	UCU	10.5	10.4	10.9	13.1	11.1	16.3	/	UCC	0.17	0.2	0.29	UCC
*His	CAC	10.8	10.7	8.8	13.9	13.9	14.2	0.4	CAU	0.14	0.19	0.31	CAU
Leu	UUA	10.5	10.9	15	6.13	4.64	2.73	0.25	UUG	0.25	0.29	0.27	UUG
§Leu	UUG	11.3	11.5	12.9	6.63	5.72	4.27	0.25	UUA	0.43	0.49	0.66	/
	UUG									0.25	0.29	0.27	UUA
*His	CAU	11.3	11.6	12.5	9.23	8.11	6.78	0.4	CAC	0.14	0.19	0.31	CAC
#Val	GUA	12.1	11.6	11.9	15.9	18.7	22.3	1.05	GUU.GUG	0.86	0.78	1.45	GUU,GUG
*#Lys	AAG	11.9	12	13.1	12.1	13.7	17.2	1.0	AAA	0.43	0.52	0.74	AAA
§Thr	ACG	12.5	12.7	13.8	7.53	6.21	4.17	minor	ACA	0.12	0.15	0.22	/
	ACG									0.2	0.23	0.49	ACA,ACU
*#Asn	AAU	16.3	16.3	22.8	9.79	7.79	5.61	0.6	AAC	0.27	0.31	0.52	AAC

a. Low-usage sense codons of *E. coli* were selected and include: all the codons at a cut-off of < 1. 28% frequency calculated from GenBank release #69, and all the codons at a cut off < 0.869% real-time frequency when *E. coli* was at the growth rate of 1.07 doublings/hour. The cut-offs are the lowest frequency values among those of non-degenerate codons (Met and Trp) and the optimal codons for amino acids with more than 2 degenerate codons (Leu, Ile, Val, Ser, Pro, Thr, Ala, Arg and Gly).

b. The data are codon usage frequencies calculated from the sequence data of GenBank release #63, 69 or 147 (refer to Table 1) or calculated by Dong et al, 1996 from 140 proteins coding frames when *E. coli* was at different growth rate of 0.4 doublings/hour (refer to Table 2).

c. The relative amount is the amount relative to that of tRNA$^{Leu1(CUG)}$ that is normalized to 1.0.

a. Taken and adapted from Ikemura Ikemura (1981a,b, 1985) (refer to Table 3).

b. The other synonymous codons that are also recognized by the same tRNA.

c. Taken and adapted from Dong et al, 1996 (refer to Table 4).

d. The tRNA^{ILe2} co-migrated with tRNA^{ILe1} on 2-D PAGE. The latter recognizes AUU and AUC while the former recognizes AUA, and the relative content for different growth rate is the sum of both tRNAs. The tRNA$^{Gly1(GGG)}$ also co-migrated with tRNA$^{Gly2\ (GGA,GGG)}$ on 2-D PAGE, and the relative content for different growth rate is similarly the sum of both tRNAs.

* Cys, His, Lys and Asn all have 2 synonymous codons that are recognized by one single tRNA.

§ Codons ACU and ACG of Thr are decoded by 3 and 2 tRNAs, respectively. Codon UUG of Leu are also decoded by 2 tRNAs.

# The relative amounts for the single tRNAs which decodes GUA (Val), AAG (Lys) and AAU (Asn) are high (> 0.6 according to Ikemura (1985) or > = 0.27 according to Dong et al, 1996).

needs further studies. However, the rare sense codons in Group I, especially Arg rare codons AGG and AGA, have been extensively studied, and most effects of RCs and RCCs as well as their underlying mechanisms are obtained from studies of this group of rare sense codons (Hackett and Reeves, 1983; Pedersen, 1984; Misra and Reeves, 1985; Fang et al, 1986; Garcia et al, 1986; Pohlner et al, 1986; Harms and Umbarger, 1987; Chen et al, 1990a,b, 1991; Gurskii et al, 1992a, b; Ivanov et al, 1992; Kane et al, 1992; Gursky et al, 1994; Hua et al, 1994, 1996; Vilbois et al, 1994; Curran, 1995; Del, Jr. et al, 1995; Kane, 1995; Bouquin et al, 1996; Calderone et al, 1996; Major et al, 1996; Saraffova et al, 1996; Zahn and Landy, 1996; Zahn, 1996; Babic et al, 1997; Ivanov et al, 1997; Schwartz and Curran, 1997; Tsai and Curran, 1998; Wakagi et al, 1998; Jiang et al, 1999; Imamura et al, 1999; Roche and Sauer, 1999; Sauer and Nygaard, 1999; Kleber-Janke et al, 2000; Zdanovsky and Zdanovskaia, 2000; Acosta-Rivero et al, 2002; Hayes et al, 2002; Kapust et al, 2002; Laine et al, 2002; Park et al, 2002; McNulty et al, 2003; Olivares-Trejo et al, 2003; Tan et al, 2003; Chen et al, 2004; Sakamoto et al, 2004; Shu et al, 2004; Gurvich et al, 2005).

Codon optimization (a kind of nucleotide substitution which replaces the rare codons in a gene by synonymous optimal or other major codons) and rare tRNA supplementation (co-expression of rare-tRNA genes) are the 2 strategies to overcome the expression problems caused by rare sense codons or study the underlying mechanisms. In order to highly express a foreign gene in bacterium *E. coli*, either one or both of the 2 strategies may be adopted. Because mutant strains such as *Rosetta 2(DE3) of E. coli* (Chen et al, 2004) are commercially available for rare tRNA supplementation, it is recommended to first try this strategy when a foreign gene cannot be expressed to satisfaction in regular expression host such as *BL21(DE3)*. Further, if rare tRNA supplementation cannot correct the expression problem(s), codon optimization to replace some or all of the *E. coli* RCs or their RCCs in a foreign gene is likely to be a must. The *E. coli* RCs that should be considered in codon optimization, based on the so far reports, at least include all the above 7 Group I RCs and probably the 6 Group IIa RCs.

References

Acosta-Rivero N, Sanchez JC and Morales J (**2002)** Improvement of human interferon HUIFNα2 and HCV core protein expression levels in Escherichia coli but not of HUIFNα8 by using the tRNA (AGA/AGG). **Biochem Biophys Res Commun** 296, 1303-1309.

Babic S, Hunter CN, Rakhlin NJ, Simons RW and Phillips-Jones MK (**1997)** Molecular characterisation of the pifC gene encoding translation initiation factor 3, which is required for normal photosynthetic complex formation in Rhodobacter sphaeroides NCIB 8253. **Eur J Biochem** 249, 564-575.

Bouquin N, Chen MX, Kim S, Vannier F, Bernard S, Holland IB and Seror SJ (**1996)** Characterization of an Escherichia coli mutant, feeA, displaying resistance to the calmodulin inhibitor 48/80 and reduced expression of the rare tRNA3Leu. **Mol Microbiol** 20, 853-865.

Bulmer M (**1987)** Coevolution of codon usage and transfer RNA abundance. **Nature, 325** 728-730.

Calderone TL, Stevens RD and Oas TG (**1996)** High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in Escherichia coli. **J Mol Biol** 262, 407-412.

Chen D, Duggan C, Ganley JP, Kooragayala LM, Reden TB, Texada DE and Langford MP (**2004)** Expression of enterovirus 70 capsid protein VP1 in Escherichia coli. **Protein Expr Purif** 37, 426-433.

Chen GF and Inouye M (**1990a)** Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the Escherichia coli genes. **Nucleic Acids Res** 18, 1465-1473.

Chen KS, Peters TC and Walker JR (**1990b)** A minor arginine tRNA mutant limits translation preferentially of a protein dependent on the cognate codon. **J Bacteriol** 172, 2504-2510.

Chen MX, Bouquin N, Norris V, Casaregola S, Seror SJ and Holland IB (**1991)** A single base change in the acceptor stem of tRNA (3Leu) confers resistance upon Escherichia coli to the calmodulin inhibitor, 48/80. **EMBO J** 10, 3113-3122.

Choi AH, Basu M, McNeal MM, Bean JA, Clements JD and Ward RL (**2004)** Intranasal administration of an Escherichia coli-expressed codon-optimized rotavirus VP6 protein induces protection in mice. **Protein Expr Purif** 38, 205-216.

Crick FH (**1966)** Codon--anticodon pairing: the wobble hypothesis. **J Mol Biol** 19, 548-555.

Curran JF (**1995)** Decoding with the A:I wobble pair is inefficient. **Nucleic Acids Res** 23, 683-688.

Del TB Jr, Ward JM, Hodgson J, Gershater CJ, Edwards H, Wysocki LA, Watson FA, Sathe G and Kane JF (**1995)** Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli. **J Bacteriol** 177, 7086-7091.

Dong H, Nilsson L and Kurland CG (**1996)** Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. **J Mol Biol** 260, 649-663.

Emilsson V and Kurland CG (**1990)** Growth rate dependence of transfer RNA abundance in Escherichia coli. **EMBO J** 9, 4359-4366.

Emilsson V, Naslund AK and Kurland CG (**1993**) Growth-rate-dependent accumulation of twelve tRNA species in Escherichia coli. **J Mol Biol** 230, 483-491.

Fang GH, Kenigsberg P, Axley MJ, Nuell M and Hager LP (**1986**) Cloning and sequencing of chloroperoxidase cDNA. **Nucleic Acids Res** 14, 8061-8071.

Flick K, Ahuja S, Chene A, Bejarano MT and Chen Q (**2004**) Optimized expression of Plasmodium falciparum erythrocyte membrane protein 1 domains in Escherichia coli. **Malar J** 3, 50.

Garcia GM, Mar PK, Mullin DA, Walker JR and Prather NE (**1986**) The E. coli dnaY gene encodes an arginine transfer RNA. **Cell** 45, 453-459.

Garel JP (**1974**) Functional adaptation of tRNA population. **J Theor Biol** 43, 211-225.

Garel JP, Chavancy G, Chevallier A, Fournier A, Marbaix G and Huez G (**1981**) [tRNA adaptation and the optimization of translation]. **Reprod Nutr Dev** 21, 177-183.

Gold L (**1990**) Expression of heterologous proteins in Escherichia coli. **Methods Enzymol** 185, 11-14.

Gouy M and Gautier C (**1982**) Codon usage in bacteria: correlation with gene expressivity. **Nucleic Acids Res** 10, 7055-7074.

Grantham R, Gautier C and Gouy M (**1980a**) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. **Nucleic Acids Res** 8, 1893-1912.

Grantham R, Gautier C, Gouy M, Jacobzone M and Mercier R (**1981**) Codon catalog usage is a genome strategy modulated for gene expressivity. **Nucleic Acids Res** 9, 43-74.

Grantham R, Gautier C, Gouy M, Mercier R and Pave A (**1980b**) Codon catalog usage and the genome hypothesis. **Nucleic Acids Res** 8, 49-62.

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.

Grosjean H, Sankoff D, Jou WM, Fiers W and Cedergren RJ (**1978**) Bacteriophage MS2 RNA: a correlation between the stability of the codon: anticodon interaction and the choice of code words. **J Mol Evol** 12, 113-119.

Gurskii I, Marimont NI and Bibilashvili RS (**1992a**) [The effect of intracellular concentrations of tRNA, corresponding to the rare arginine codons AGG and AGA, on the gene expression in Escherichia coli]. **Mol Biol (Mosk)** 26, 1080-7.

Gurskii I, Marimont NI, Shevelev AI, Iuzhakov AA and Bibilashvili RS (**1992b**) [Rare codons and gene expression in Escherichia coli]. **Mol Biol (Mosk)** 26, 1063-79.

Gursky YG and Beabealashvilli RS (**1994**) The increase in gene expression induced by introduction of rare codons into the C terminus of the template. **Gene** 148, 15-21.

Gurvich OL, Baranov PV, Gesteland RF and Atkins JF (**2005**) Expression levels influence ribosomal frameshifting at the tandem rare arginine codons AGG_AGG and AGA_AGA in Escherichia coli. **J Bacteriol** 187, 4023-4032.

Gutman GA and Hatfield GW (**1989**) Nonrandom utilization of codon pairs in Escherichia coli. **Proc Natl Acad Sci U S A** 86, 3699-3703.

Hackett J and Reeves P (**1983**) Primary structure of the tolC gene that codes for an outer membrane protein of Escherichia coli K12. **Nucleic Acids Res** 11, 6487-6495.

Harms E and Umbarger HE (**1987**) Role of codon choice in the leader region of the ilvGMEDA operon of Serratia marcescens. **J Bacteriol** 169, 5668-5677.

Hayes CS, Bose B and Sauer RT (**2002**) Stop codons preceded by rare arginine codons are efficient determinants of SsrA tagging in Escherichia coli. **Proc Natl Acad Sci U S A** 99, 3440-3445.

Hodgson J (**1993**) Expression systems: a user's guide. Emphasis has shifted from the vector construct to the host organism. **Biotechnology (NY)** 11, 887-893.

Hu X, Shi Q, Yang T and Jackowski G (**1996**) Specific replacement of consecutive AGG codons results in high-level expression of human cardiac troponin T in Escherichia coli. **Protein Expr Purif** 7, 289-293.

Hua Z, Wang H, Chen D, Chen Y and Zhu D (**1994**) Enhancement of expression of human granulocyte-macrophage colony stimulating factor by argU gene product in Escherichia coli. **Biochem Mol Biol Int** 32, 537-543.

Ikemura T (**1980**) [Measurement of relative amount of E. *coli tRNAs:* codon choice in E. coli genes is largely constrained by the concentration of anticodons (author's transl)]. **Tanpakushitsu Kakusan Koso** 25, 668-678.

Ikemura T (**1981a**) 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.

Ikemura T (**1981b**) 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.

Ikemura T (**1982**) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. **J Mol Biol** 158, 573-597.

Ikemura T (**1985**) Codon usage and tRNA content in unicellular and multicellular organisms. **Mol Biol Evol** 2, 13-34.

Imamura H, Jeon B, Wakagi T and Matsuzawa H (**1999**) High level expression of Thermococcus litoralis 4-α-glucanotransferase in a soluble form in Escherichia coli with a novel expression system involving minor arginine tRNAs and GroELS. **FEBS Lett** 457, 393-396.

Irwin B, Heck JD and Hatfield GW (**1995**) Codon pair utilization biases influence translational elongation step times. **J Biol Chem** 270, 22801-22806.

Ivanov I, Alexandrova R, Dragulev B, Saraffova A and Abouhaidar MG (**1992**) Effect of tandemly repeated AGG triplets on the translation of CAT-mRNA in E. **coli FEBS Lett** 307, 173-176.

Ivanov IG, Saraffova AA and Abouhaidar MG (**1997**) Unusual effect of clusters of rare arginine (AGG) codons on the expression of human interferon α 1 gene in Escherichia coli. **Int J Biochem Cell Biol** 29, 659-666.

Jiang L, Yang Y, Chatterjee S, Seidel B, Wolf G and Yang S (**1999**) The expression of proUK in Escherichia coli: the vgb promoter replaces IPTG and coexpression of argU compensates for rare codons in a hypoxic induction model. **Biosci Biotechnol Biochem** 63, 2097-2101.

Jonasson P, Liljeqvist S, Nygren PA and Stahl S (**2002**) Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli. **Biotechnol Appl Biochem** 35, 91-105.

Kane JF (**1995)** Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. **Curr Opin Biotechnol** 6, 494-500.

Kane JF, Violand BN, Curran DF, Staten NR, Duffin KL and Bogosian G (**1992)** Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of Escherichia coli. **Nucleic Acids Res** 20, 6707-6712.

Kapust RB, Routzahn KM and Waugh DS (**2002)** Processive degradation of nascent polypeptides, triggered by tandem AGA codons, limits the accumulation of recombinant tobacco etch virus protease in Escherichia coli BL21 (DE3). **Protein Expr Purif** 24, 61-70.

Kleber-Janke T and Becker WM (**2000)** Use of modified BL21 (DE3) Escherichia coli cells for high-level expression of recombinant peanut allergens affected by poor codon usage. **Protein Expr Purif** 19, 419-424.

Konigsberg W and Godson GN (**1983)** Evidence for use of rare codons in the dnaG gene and other regulatory genes of Escherichia coli. **Proc Natl Acad Sci U S A** 80, 687-691.

Laine S, Salhi S and Rossignol JM (**2002)** Overexpression and purification of the hepatitis B e antigen precursor. **J Virol Methods** 103, 67-74.

Lipman DJ and Wilbur WJ (**1983)** Contextual constraints on synonymous codon choice. **J Mol Biol** 163, 363-376.

Ma HH, Yang L, Yang XY, Xu ZP and Li BL (**2003)** Bacterial expression, purification, and in vitro N-myristoylation of fusion hepatitis B virus preS1 with the native-type N-terminus. **Protein Expr Purif** 27, 49-54.

Major LL, Poole ES, Dalphin ME, Mannering SA and Tate WP (**1996)** Is the in-frame termination signal of the Escherichia coli release factor-2 frameshift site weakened by a particularly poor context? . **Nucleic Acids Res** 24, 2673-2678.

Makrides SC (**1996)** Strategies for achieving high-level expression of genes in Escherichia coli. **Microbiol Rev** 60, 512-538.

McNulty DE, Claffee BA, Huddleston MJ and Kane JF (**2003)** Mistranslational errors associated with the rare arginine codon CGG in Escherichia coli. **Protein Expr Purif** 27, 365-374.

Misra R and Reeves P (**1985)** Intermediates in the synthesis of TolC protein include an incomplete peptide stalled at a rare Arg codon. **Eur J Biochem** 152, 151-155.

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.

Nussinov R (**1981)** Eukaryotic dinucleotide preference rules and their implications for degenerate codon usage. **J Mol Biol** 149, 125-131.

Olins PO and Lee SC (**1993)** Recent advances in heterologous gene expression in Escherichia coli. **Curr Opin Biotechnol** 4, 520-525.

Olivares-Trejo JJ, Bueno-Martinez JG, Guarneros G and Hernandez-Sanchez J (**2003)** The pair of arginine codons AGA AGG close to the initiation codon of the lambda int gene inhibits cell growth and protein synthesis by accumulating peptidyl-tRNAArg4. **Mol Microbiol** 49, 1043-1049.

Park SJ, Lee SK and Lee BJ (**2002)** Effect of tandem rare codon substitution and vector-host combinations on the expression of the EBV gp110 C-terminal domain in Escherichia coli. **Protein Expr Purif** 24, 470-480.

Pedersen S (**1984)** Escherichia coli ribosomes translate in vivo with variable rate. **EMBO J** 3, 2895-2898.

Pedersen S, Bloch PL, Reeh S and Neidhardt FC (**1978)** Patterns of protein synthesis in E. coli: a catalog of the amount of 140 individual proteins at different growth rates **Cell** 14, 179-190.

Pohlner J, Meyer TF, Jalajakumari MB and Manning PA (**1986)** Nucleotide sequence of ompV, the gene for a major Vibrio cholerae outer membrane protein. **Mol Gen Genet** 205, 494-500.

Roche ED and Sauer RT (**1999)** SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. **EMBO J** 18, 4579-4589.

Sakamoto K, Ishimaru S, Kobayashi T, Walker JR and Yokoyama S (**2004)** The Escherichia coli argU10 (Ts) phenotype is caused by a reduction in the cellular level of the argU tRNA for the rare codons AGA and AGG. **J Bacteriol** 186, 5899-5905.

Saraffova A, Maximova V, Ivanov IG and Abouhaidar MG (**1996)** Comparative study on the effect of signal peptide codons and arginine codons on the expression of human interferon-α 1 gene in Escherichia coli. **J Interferon Cytokine Res** 16, 745-749.

Sauer J and Nygaard P (**1999)** Expression of the Methanobacterium thermoautotrophicum hpt gene, encoding hypoxanthine (Guanine) phosphoribosyltransferase, in Escherichia coli. **J Bacteriol** 181, 1958-1962.

Schwartz R and Curran JF (**1997)** Analyses of frameshifting at UUU-pyrimidine sites. **Nucleic Acids Res** 25, 2005-2011.

Sharp PM and Li WH (**1986)** Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons. **Nucleic Acids Res** 14, 7737-7749.

Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH and Wright F (**1988)** Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. **Nucleic Acids Res** 16, 8207-8211.

Shu P, Dai H, Mandecki W and Goldman E (**2004)** CCC CGA is a weak translational recoding site in Escherichia coli. **Gene** 343, 127-132.

Sorensen HP and Mortensen KK (**2005a)** Advanced genetic strategies for recombinant protein expression in Escherichia coli. **J Biotechnol** 115, 113-128.

Sorensen HP and Mortensen KK (**2005b)** Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. **Microb Cell Fact** 4, 1.

Sorensen MA, Kurland CG and Pedersen S (**1989)** Codon usage determines translation rate in Escherichia coli. **J Mol Biol** 207, 365-377.

Tan WS, Dyson MR and Murray K (**2003)** Hepatitis B virus core antigen: enhancement of its production in Escherichia coli, and interaction of the core particles with the viral surface antigen. **Biol Chem** 384, 363-371.

Tsai F and Curran JF (**1998)** tRNA (2Gln) mutants that translate the CGA arginine codon as glutamine in Escherichia coli. **RNA** 4, 1514-1522.

Vilbois F, Caspers P, da Prada M, Lang G, Karrer C, Lahm HW and Cesura AM (**1994)** Mass spectrometric analysis of human soluble catechol O-methyltransferase expressed in Escherichia coli. Identification of a product of ribosomal frameshifting and of reactive cysteines involved in S-adenosyl-L-methionine binding **Eur J Biochem** 222, 377-386.

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

Wakagi T, Oshima T, Imamura H and Matsuzawa H (**1998)** Cloning of the gene for inorganic pyrophosphatase from a thermoacidophilic archaeon, Sulfolobus sp. strain 7, and overproduction of the enzyme by coexpression of tRNA for arginine rare codon **Biosci Biotechnol Biochem** 62, 2408-2414.

Zahn K (**1996)** Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. **J Bacteriol** 178, 2926-2933.

Zahn K and Landy A (**1996)** Modulation of lambda integrase synthesis by rare arginine tRNA. **Mol Microbiol** 21, 69-76.

Zdanovsky AG and Zdanovskaia MV (**2000)** Simple and efficient method for heterologous expression of clostridial proteins. **Appl Environ Microbiol** 66, 3166-3173.

Zhang SP, Zubay G and Goldman E (**1991)** Low-usage codons in Escherichia coli, yeast, fruit fly and primates. **Gene** 105, 61-72.

Zhou Z, Schnake P, Xiao L and Lal AA (**2004)** Enhanced expression of a recombinant malaria candidate vaccine in Escherichia coli by codon optimization. **Protein Expr Purif** 34, 87-94.

Dequan Chen