Gene, 18 (1982) 199–209
Elsevier Biomedical Press

Review

Preferential codon usage in prokaryotic genes: the optimal codon–anticodon interaction energy and the selective codon usage in efficiently expressed genes

(RNA bacteriophage; frameshift; modulating codons; efficiency of translation; isoaccepting tRNAs)

Henri Grosjean * and Walter Fiers

* *Laboratoire de Chimie Biologique, Université de Bruxelles, 67, rue des Chevaux, B-1640 Rhode St. Genèse and Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat, 35, B-9000 Ghent (Belgium)*

(Received October 28th, 1981)
(Accepted April 29th, 1982)

SUMMARY

By considering the nucleotide sequence of several highly expressed coding regions in bacteriophage MS2 and mRNAs from *Escherichia coli*, it is possible to deduce some rules which govern the selection of the most appropriate synonymous codons NNU or NNC read by tRNAs having GNN, QNN or INN as anticodon. The rules fit with the general hypothesis that an efficient in-phase translation is facilitated by proper choice of degenerate codewords promoting a codon-anticodon interaction with intermediate strength (*optimal energy*) over those with very strong or very weak interaction energy. Moreover, codons corresponding to minor tRNAs are clearly avoided in these efficiently expressed genes. These correlations are clearcut in the normal reading frame but *not* in the corresponding frameshift sequences +1 and +2. We hypothesize that both the optimization of codon-anticodon interaction energy and the adaptation of the population to codon frequency or vice versa in highly expressed mRNAs of *E. coli* are part of a strategy that optimizes the efficiency of translation. Conversely, codon usage in weakly expressed genes such as repressor genes follows exactly the opposite rules. It may be concluded that, in addition to the need for coding an amino acid sequence, the energetic consideration for codon-anticodon pairing, as well as the adaptation of codons to the tRNA population, may have been important evolutionary constraints on the selection of the optimal nucleotide sequence.

INTRODUCTION

The amount of information on nucleotide sequences of messenger RNAs or genes is now expanding extremely rapidly, due to the considerable progress being made in the techniques for primary structure determination as well as for isolating messenger RNAs or cloned genes. The interpretation of such information in terms of biological properties is, however, still an arduous task. Ultimately, we may hope to understand these data in the more general framework of the evolution and stability of the genetic material. Obviously,

Abbreviations: EF, elongation factor; IF, initiation factor; N, nucleotide; pfu, plaque-forming units. (Abbreviations of modified bases in anticodons of tRNA are listed in the legend to Table IV).

0378-1119/82/0000-0000/$02.75

during evolution, many parameters may have played a role at different levels of the genetic expression and organization (replication, transcription, maturation, translation, stability, packaging and so on); they may all have imposed different constraints on the selection of a given sequence in DNA or RNA that we may not be able to discern easily.

Many years ago, we drew attention to the strikingly non-random codon usage in the bacteriophage MS2 genome (Fiers et al., 1971; Min Jou et al., 1971; 1972). Two main ideas emerged: first, that the rare codons which correspond to minor tRNAs in *E. coli* may be used to regulate the elongation rate during translation; second, that the choice of the synonymous codons NNU and NNC which are translated by one single tRNA having an anticodon GNN, QNN or INN may depend on the stability of the codon-anticodon interaction (Grosjean et al., 1978b; cf. also review by Fiers, 1979).

A consideration of the large amount of new information on anticodon sequences (Gauss and Sprinzl, 1981), on tRNA populations in *E. coli* (Ikemura, 1981), as well as on nucleotide sequences of *E. coli* mRNAs (compiled by Grantham et al., 1981), has led us to reinvestigate and extend our previous conclusions.

The present paper shows that the rules which we initially deduced for the MS2 genome are also valid for the bacterial mRNAs and allow a better understanding of the strikingly different codon usage in strongly expressed and in weakly expressed bacterial mRNAs.

RESULTS AND DISCUSSION

(a) The optimal codon–anticodon interaction energy hypothesis

(1) Application to the MS2 RNA bacteriophage

Briefly, our original hypothesis was as follows (Grosjean et al., 1978b; Fiers and Grosjean, 1979). Upon infection, the phage message has to be translated efficiently and accurately to rapidly impose its genetic blueprint on the host cell: in the case of MS2 RNA the burst size in *E. coli* is 5000 to 10000 pfu per cell, and each virus particle contains 180 coat protein molecules (Fiers, 1979). Such an optimal translation especially of the coat gene would proceed efficiently if the codon frequency is well adapted to the tRNA population (see below); and it would proceed most smoothly if the interaction energy between codons and corresponding anticodons were rather uniform. For degenerate codons recognized by the same tRNA an intermediate strength for codon–anticodon interaction (rather than a very strong or very weak interaction) can be obtained by adapting the third letter of the codons. A selective codon usage supporting this hypothesis does indeed exist for translation of codons containing a pyrimidine as the third letter, at least in prokaryotic mRNA.

The rationale for the hypothesis is as follows: in *E. coli* both NNC and NNU codons are well known to be read by their corresponding tRNAs having an anticodon GNN or QNN (INN for the major arginine tRNA) (see Table IV). Indeed, unmodified adenosine is never present in the so-called "wobble" position of an anticodon (see Gauss and Sprinzl, 1981). We may therefore expect that the NNU codons will form a relatively weak complex with the corresponding GNN, QNN or INN anticodons. In fact, this is exactly the case: a factor of three to ten difference in the equilibrium constant as well as in the kinetic rate constant for the dissociation of the codon-anticodon complex has been measured with several RNAs (Freier and Tinoco, 1975; D. Labuda, H. Grosjean, G. Stricker and D. Pörschke, submitted for publication, and references therein). The same is true where two tRNAs have complementary anticodons (Grosjean et al., 1978a; Weissenbach and Grosjean, 1981). Therefore, during the translation process on the ribosome, codons in which the first two nucleotides NN are U and/or A should stabilize their interaction with the GNN, QNN (or INN) anticodons if a C is used in the wobble position. In the whole genome of MS2-RNA, including the lysis protein that is read in a different frame (Atkins et al., 1979a; Beremand and Blumenthal, 1979), we note that there is indeed a systematic preference for the NNC codons over NNU for phenylalanine (codon $UU_C^U : \frac{21}{32}$), isoleucine ($AU_C^U : \frac{12}{28}$), tyrosine ($UA_C^U : \frac{10}{33}$) and asparagine ($AA_C^U : \frac{19}{28}$).

TABLE I

Codon usage in bacteriophage MS2

Total number of each of the 64 codons used in the four coding regions of MS2-RNA are given in the conventional genetic code grid. **A** denotes A-protein or maturase (394 codons), **C** coat (131 codons), **R** replicase (545 codons) and **L** lysis protein (75 codons). Σ is the sum (1145 codons). The lysis protein gene overlaps the coat and replicase gene; it starts near the end of the coat protein, in frame +1 as compared to the coat and the replicase genes (Atkins et al., 1979a; Beremand and Blumenthal, 1979). All numbers come from analysis of the complete MS2-RNA sequence (Fiers et al., 1976). The boxes correspond to codon pairs where the choice between the degenerate codons is expected to be most dependent on the optimization of the codon-anticodon interaction energy effect (observed preference shown in heavy type). Arrowheads indicate putative modulator codons not used in the coat protein gene.

		U						C						A						G					
		A	C	R	L	Σ		A	C	R	L	Σ		A	C	R	L	Σ		A	C	R	L	Σ	
U	Phe	6	1	12	2	21	Ser	5	3	7	1	16	Tyr	4	0	5	1	10	Cys	0	1	5	1	7	U
		10	**3**	**16**	**3**	**32**		6	2	12	0	20		**12**	**4**	**16**	1	**33**		3	1	2	0	6	C
	Leu	8	1	8	2	19		8	2	6	1	17	ochre	0	1	0	1	2	opal	0	0	0	0	0	A
		6	0	5	2	13		10	2	10	3	25	amber	1	0	1	0	2	Trp	12	2	9	0	23	G
C	Leu	6	2	7	3	18	Pro	5	2	**10**	1	**18**	His	2	0	4	1	7	Arg	7	3	11	0	21	U
		9	2	15	2	28		5	1	4	0	10		3	0	6	0	9		6	1	13	2	22	C
		5	2	8	1	16		4	2	3	2	11	Gln	9	1	7	6	23		6	▶0	4	3	13	A
		2	0	7	2	11		3	1	9	1	14		9	5	8	3	25		3	▶0	8	0	11	G
A	Ile	1	4	7	0	12	Thr	11	4	4	5	24	Asn	2	4	11	2	19	Ser	4	0	4	1	9	U
		8	4	**13**	**3**	**28**		5	4	12	2	23		**15**	**6**	7	0	**28**		3	4	9	0	16	C
		7	▶0	12	0	19		5	0	8	1	14	Lys	5	5	9	2	21	Arg	3	▶0	4	2	9	A
	Met	7	3	10	1	21		6	1	7	1	15		9	1	16	0	26		4	▶0	2	0	12	G
G	Val	8	4	9	0	21	Ala	6	**5**	**15**	0	**26**	Asp	8	1	19	1	29	Gly	**15**	3	**19**	0	**37**	U
		7	4	10	0	21		12	2	7	0	21		5	3	14	0	22		6	3	7	0	16	C
		7	3	6	1	17		7	6	8	1	22	Glu	5	2	9	2	18		2	2	8	0	12	A
		9	3	7	2	17		10	1	12	2	25		12	3	13	1	29		5	1	10	0	16	G

Conversely, if the first two nucleotides NN of the codon are C and/or G, a preference would exist for U as a third letter to avoid too sticky a codon–anticodon interaction. Indeed, the NNU is systematically preferred over NNC for proline (codon $CC_C^{U:18}_{10}$), alanine ($GC_C^{U:26}_{21}$) and glycine ($GG_C^{U:37}_{16}$) albeit not for arginine ($CG_C^{U:21}_{22}$). Apart from the fact that reducing the sample size decreases the statistical significance of these numbers this trend applies satisfactorily to each individual gene of MS2 (see Table I).

Fitch (1976) had already examined the problem of selection of a wobble codon in the MS2 genome based on the strength of codon-anticodon pairing. He concluded, however, that the selection of degenerate codons was against wobble because there were more NNC (142) than NNU (93) codons used in the whole coding parts of the MS2 genome. Our interpretation (Grosjean et al., 1978b) of the problem, however, was essentially based on the distinction we made amongst the intrinsically weak (AU-rich) code words and the intrinsically strong (GC-rich) code words. Mixing of the two classes of codons, of course, completely masked the evidence that the choice of a G.U or G.C interaction of the wobble may play a role in *modulating* the strength of pairing between codon and anticodon; this modulation means that the codon-anticodon interaction should be neither too loose, nor too sticky.

This trend may not necessarily be seen in all other bacteriophage genomes (RNA or DNA,

single- or double-stranded), either because they are expressed in their host with less efficiency than MS2 RNA or because there are other important evolutionary constraints that have prevented optimization of codon-anticodon pairing energy. For example, constraints at the level of the replication process, on packaging and/or on the metabolic stability of the bacteriophage genome itself.

(2) Application to mRNAs that code for highly expressed proteins in Escherichia coli

In recent years, sequence information has become available on other proteins that are synthesized in abundant amounts in *E. coli*, such as ribosomal proteins (Post et al., 1979, 1980; Post and Nomura, 1980), the *recA* protein (abundant after induction; Horii et al., 1980; Sancar et al.,

TABLE II

Codon usage in highly expressed and weakly expressed *E. coli* genes

Total number of each of the codons used in 23 mRNAs corresponding to abundant proteins in *E. coli* (columns 1 to 4). In the last column (rep), we present the cumulative data for four repressor proteins that are expressed very weakly in *E. coli*. Symbol pol stands for RNA polymerases (total codons assigned: 2274) and includes the complete gene for *rpoD* (Burton et al., 1981), *rpoB* (Ovchinnikov et al., 1980, cited in Burton et al., 1981), part of the gene for *rpoC* (Post et al., 1979; Gurevich et al., 1980) and *rpoA* (Post and Nomura, 1979). Symbol r-pro stands for ribosomal proteins (total codons assigned 1108); it includes the complete genes for protein S12, L11, L1, L10 and L7/12, as well as portions of the gene for protein S4, S7, S11, S13, L14 and L16 (see Post and Nomura, 1980). Symbol fac stands for factors (total codons 1019); it includes the *tufA* gene (Yokota et al., 1980) but not the *tufB* gene which is almost the same as *tufA*, EF-G factor (Post and Nomura, 1980; Yokota et al., 1980), IF3 (Fayat, G., Sacerdot, C. and Blanquet, S. personal communication). Also included is the *recA* gene (Horii et al., 1980). Symbol o-mbr stands for outer membrane proteins (458 codons assigned), including the lipoprotein (Nakamura et al., 1980), *lamB* protein (Hedgepeth et al., 1980) and *ompA* protein II (Beck and Bremer, 1980). Symbol rep stands for repressor protein (915 codons assigned); it includes the *lacI* gene (Farabaugh, 1978), the *trpR* gene (Singleton et al., 1980), the *araC* gene (Miyada et al., 1980) and the Tn3-coded repressor (Heffron et al., 1979). The values indicate the frequency of occurrence of each codon for the different groups of mRNAs. The values in bold-face letters show clear-cut degenerate codon preferences as discussed in this paper. Arrows point out the putative modulator codons corresponding to minor (or weakly interacting) tRNAs (cf. Table IV).

		U						C						A						G						
			r-pol	r-pro	r-fac	o-mbr	rep		r-pol	r-pro	r-fac	o-mbr	rep		r-pol	r-pro	r-fac	o-mbr	rep		r-pol	r-pro	r-fac	o-mbr	rep	
U	Phe		21	9	6	2	**20**	Ser	37	24	17	8	7	Tyr	21	4	5	3	**13**	Cys	5	1	5	1	4	U
			50	**17**	**25**	**8**	7		42	21	10	11	9		**42**	**13**	**18**	**16**	9		10	6	3	2	4	C
	Leu		5	4	2	1	14		3	1	2	0	10	ochre	2	7	3	2	1	opal	0	0	0	0	3	A
			9	3	3	1	14		9	2	1	0	15	amber	0	0	0	0	0	Trp	10	4	5	5	9	G
C	Leu		14	4	6	1	10	Pro	**11**	**4**	**4**	**2**	1	His	7	4	4	2	14	Arg	**101**	**46**	**42**	**14**	14	U
			23	3	7	0	13		0	1	1	0	**8**		31	9	15	4	9		55	24	14	5	**29**	C
			→ 3	0	0	0	3		13	6	3	3	5	Gln	16	8	5	2	25		→ 2	0	1	0	11	A
			162	67	57	32	49		64	32	32	15	20		83	26	31	21	33		→ 0	1	0	0	10	G
A	Ile		35	16	11	2	**26**	Thr	24	31	21	15	6	Asn	5	4	3	1	**19**	Ser	4	4	2	0	8	U
			115	**37**	**67**	**17**	19		60	21	30	11	22		**70**	**32**	**26**	**24**	14		26	7	9	6	18	C
			→ 0	1	1	0	4		4	3	5	1	3	Lys	92	77	51	22	22	Arg	→ 1	1	1	0	2	A
	Met		69	24	27	10	22		16	2	8	1	11		44	28	23	5	10		→ 1	0	0	0	5	G
G	Val		66	47	35	22	15	Ala	**32**	**75**	**23**	**32**	11	Asp	72	14	17	9	30	Gly	**89**	**44**	**47**	**27**	15	U
			25	8	5	2	19		22	13	10	2	**31**		91	31	40	22	22		58	35	44	17	**17**	C
			39	42	17	10	8		33	42	21	17	14	Glu	167	58	69	9	30		→ 0	1	3	0	6	A
			41	17	19	3	25		56	27	31	6	34		58	15	22	4	21		→ 8	0	5	0	18	G

1980), the *tufA* and *tufB*-coded proteins corresponding to elongation factor EF-TU (Yokota et al., 1980; An and Friesen, 1980), parts of the elongation factor EF-G (Yokota et al., 1980; Post and Nomura, 1980), the initiation factor IF3 (Fayat, G., Sacerdot, C. and Blanquet, S. personal communication), the subunits α, β, β' and σ for RNA polymerase (Post and Nomura, 1979; Gurevich et al., 1980; Burton et al., 1981) as well as three of the most abundant proteins of the outer membrane (Nakamura et al., 1980; Hedgepeth et al., 1980; Beck and Bremer, 1980). Altogether these proteins account for about 5250 codons (24 different abundant proteins). Each of these proteins exists in *E. coli* in as many as 10^3 to 7×10^5 copies per cell (see references in Gouy and Grantham, 1980; DiRienzo et al., 1978), but only one corresponding structural gene is present in the *E. coli* chromosome. So, clearly, these genes must be very efficiently transcribed as well as translated and this was shown in more detail for the EF-TU gene of *E. coli* by Young and Furano (1981). As shown in Table II, all the data follow basically the same trend as originally observed for MS2-RNA, namely preference for NNC in the case of phenylalanine, isoleucine, tyrosine and asparagine, and preference for NNU in the case of proline, alanine, arginine and glycine. These data clearly reinforce and generalize our previous hypothesis and indicate that it applies to many genes other than the MS2 genome.

We have also previously noted (Fiers and Grosjean, 1979), that a gene for a protein of which only a few molecules are needed per cell viz. the *lac* repressor protein I, follows exactly the opposite pattern. The same appears to be true for three other regulatory protein genes that are expressed at a level of only a few copies per *E. coli* cell, i.e. the repressor for the tryptophan operon (Rose and Yanofsky, 1974; Singleton et al., 1980), the regulatory protein *araC* (Miyada et al., 1980), and the specific repressor carried by the Tn*3* transposon (Heffron et al., 1979). The pooled codon usage for these four repressor proteins is also presented in Table II. It is interesting to note that the codon usage for all amino acids is more random than in the case of the mRNAs corresponding to abundant proteins; clearly specific codons are never completely avoided. These data suggest that the codon usage is much less restrictive in weakly expressed genes than in highly expressed ones and points out the importance of selecting an appropriate class of mRNAs to make an evaluation of selective codon usage.

A systematic investigation of this problem was performed using computer facilities to study a sample of 13 "highly expressed" and 16 "weakly expressed" genes from a total of 29 bacterial messengers with sequences that were known at the time (Grantham et al., 1981). Basically, the results confirm our original tenet; furthermore, the data enable them to extend it to the observation that in highly expressed mRNAs the NNC codons are also significantly preferred for the amino acids histidine (codon CA^U_C), aspartic acid (GA^U_C), cysteine (UG^U_C) and serine (AG^U_C), (see Table III).

Recently Wain-Hobson et al. (1981) have analysed codon usage in several mRNAs or genes from various origins including MS2 RNA, by statistical methods. The essence of this approach was to compare the preferential use of degenerate codons in the normal reading frame as well as in the non-coding frame of a given genetic message. This should make it possible to assess whether or not a preference for a triplet is specific to the translation process. With such a criterion, Wain-Hobson et al. (1981) claimed that the codon usage does not seem to be correlated with an optimum codon–anticodon interaction strength.

We feel that their conclusion is invalid for three main reasons. First, they have limited their comments to the analysis of the MS2 coat protein gene which contains only 130 codons. Second, when comparing sequences of mRNAs they have not taken into account the potential level of expression for an RNA molecule, as we stressed above. Third, it is not evident that the non-coding frames constitute the best available control for comparison of the codon usage in the normal reading frame. Evolutionary constraints on the optimization of a codon-anticodon strength may apply inside the codon itself as well as in its immediate context. Indeed, it is now well documented that the nucleotide immediately adjacent to a codon, especially at its 3′ side, modulates the strength of the codon-anticodon interaction (for a review see Buckingham and Grosjean, 1982). By a mechanism which is still unknown it also appears that

TABLE III

Comparison of codon usage in strongly and moderately to weakly expressed genes in *E. coli*

First (strong) column: total number of each of the 64 codons used in 24 mRNAs corresponding to abundant proteins, i.e. highly expressed genes in *E. coli*. The list of proteins is the same as in Table II except that the codon usage in *tufB* gene is also included (An and Friesen, 1980). Total number of codons assigned = 5253. Second (weak) column: total number of the 64 codons used in 18 mRNAs corresponding to non-abundant proteins in *E. coli*. The list includes the four repressor protein genes referred to in Table II as well as the genes for *lac* permease (Buchel et al., 1980) part of *ilvE* and *ilvG* genes (Lawther et al., 1979), the genes for *trpA, B, C, (G), D* and *E* (Nichols and Yanofsky, 1979; Crawford et al., 1980; Christie and Platt, 1980; Nichols et al., 1980; 1981), a gene of unknown function preceding the *ompA* gene (Beck and Bremer, 1980), the gene for dihydrofolate reductase (Smith and Calvo, 1980), the chloramphenicol acetyl transferase carried in transposon Tn*9* (Alton and Vapnek, 1979), the transposase and β-lactamase on transposon Tn*3* (Heffron et al., 1979), part of protein *Int* that precedes an attachment site for bacteriophage λ (Csordas-Toth et al., 1979). Here the list is not exhaustive but was limited to give a sample of total codons assigned (5231) that could be compared to the list of abundant proteins above. Preferential usage of degenerate codons as discussed in the text are given in heavy type. Boxes correspond to codon pairs where the choice between the degenerate codons is expected to be most dependent on the optimization of the codon–anticodon interaction energy effect. Arrows indicate the putative modulator codons corresponding to minor (or weakly interacting) tRNAs in *E. coli* (see also Table IV).

	U			C			A			G		
	STRONG	WEAK		STRONG	WEAK		STRONG	WEAK		STRONG	WEAK	
U	Phe 39	**151**	U	Ser **93**	36	U	Tyr 34	**96**	U	Cys 13	34	U
	Phe **113**	102	C	87	**49**	C	**98**	65	C	**23**	**39**	C
	Leu 12	71	A	6	37	A	ochre		A	opal		A
	Leu 16	64	G	12	62	G	amber		G	Trp 25	66	G
C	Leu 26	**73**	U	Pro **21**	29	U	His 19	**95**	U	**223**	99	U
	33	69	C	2	**46**	C	**75**	59	C	101	**133**	C
	→ 3	22	A	26	45	A	Gln 38	90	A	Arg → 3	27	A
	345	294	G	162	101	G	169	166	G	→ 1	42	G
A	Ile 67	**156**	U	Thr 103	48	U	Asn 13	**101**	U	Ser 10	56	U
	Ile **262**	118	C	**137**	**119**	C	**159**	98	C	**49**	**61**	C
	→ 2	27	A	15	32	A	Lys 259	163	A	Arg → 3	28	A
	Met 140	130	G	28	76	G	106	44	G	→ 1	17	G
G	Val **192**	**108**	U	Ala **173**	87	U	Asp 116	**183**	U	**226**	124	U
	Val 41	66	C	48	**178**	C	**204**	106	C	174	**140**	C
	119	48	A	119	107	A	Glu 333	210	A	→ 4	42	A
	83	123	G	129	149	G	106	98	G	→ 14	66	G

certain sequences of mRNA are prone to ribosome frameshift during translation (see Atkins et al., 1972; 1979b; Gallant and Foley, 1979; Fox and Weiss-Brummer, 1980; Kastelein et al., 1981); it might well be therefore, that a particular selection of preceding codons in a given mRNA is also important for an in phase translation process in order to ensure its efficiency and its accuracy, a situation that should certainly bias the codon usage in the "frame-shift triplet" as well as in the normal reading frame (see below).

(b) Modulating codons

(1) Role of the tRNA population

We now consider the selection between degenerate codons recognized by different iso-accepting tRNAs. We have already mentioned that some of the codons not used in the coat protein gene of MS2 could have a modulating role, i.e. that they regulate the rate of the translation process, and therefore are avoided in genes with high expression efficiencies (Fiers et al., 1971; Min Jou et al.,

1972). The codons AUA for isoleucine and CGG/AGA/AGG for arginine are especially strong candidates for this role, since the cognate tRNAs are very minor species in *E. coli* (reviewed in Fiers, 1979; Post et al., 1980). From a consideration of the results in Tables II and III, we may now extend this list of putative modulator codons in *E. coli* to CUA for leucine, CGA for arginine (inefficiently recognized by the ICG-containing isoaccepting tRNA) and GGA (possibly also GGG) for glycine. These codons are used in the MS2-A protein and RNA replicase genes, which are still very well translated. But it may be noted that a slight reduction of the elongation rate, for example, a reduction of 10% due to the presence of some modulating code words, may be difficult to reveal experimentally, yet may fulfill a function in fine tuning of the expression frequency of the different phage genes. It can be seen in Tables II and III that the aforementioned codons are not used at all or are very rarely used in *E. coli* genes for abundant proteins, but are used much more in weakly expressed bacterial genes. They are also remarkably low in the ϕX174 genome (Sanger et al., 1977). Moreover, it is possible that the nucleotide context around the codon influences the potential modulation role by reduction or enhancement; the same codon in a different context may not reduce the efficiency of translation to the same extent.

Alternatively, it is conceivable that the rate of elongation is modulated because of a relatively weak, or shortlived, interaction of a charged tRNA, however abundant, with certain codons at specific locations in the mRNA. Indeed, a slight modification of the structure of the tRNA anticodon such as a deficiency in hypermodification may also modulate its interaction with the different codons to which it corresponds (Weiss, 1973; Eisenberg et al., 1979; Weissenbach and Grosjean, 1981; Vacher et al., 1981). The differential utilization of isoaccepting tRNAs (hypermodified or not) and the corresponding codons may depend on the messenger considered, and it can also change with the physiological state of the cells.

Recently, Ikemura (1981) reported a comprehensive study in which a clear correlation is demonstrated between the relative abundance of different isoaccepting tRNAs in *E. coli*, and the usage of the corresponding codons in abundant protein genes. This clearly fits with the original proposition of Post et al. (1979) who pointed out that the synonymous codons preferred in *E. coli* ribosomal protein genes are those recognized by the most abundant tRNAs present in the cell, a proposition that seems to apply to all abundant gene products sequenced up to now.

However, the correlation of cellular tRNA level with codon frequency might not be so simple. In a recent review, Chavancy and Garel (1981) pointed out that the optimization effect of a balanced tRNA population on the average translation rate is mainly an overall effect. This effect is different from the specific regulatory or modulatory role of certain rare tRNAs discussed above. The interesting idea was that optimizing conditions of translation efficiency implies both a maximization of the average translation rate of a given mRNA, and a minimization of the frequency of errors (missense and frameshift) which is also certainly an essential factor in the selection of an optimal mRNA sequence during evolution. There is now experimental evidence that, in vivo, extreme imbalance in aminoacyl-tRNA species produced either by amino acid starvation or by a selective inhibition of specific aminoacylation reactions leads to increased misincorporation of amino acids and frameshifting (reviewed by Cozzone, 1980; also Gallant, 1979; Roth, 1981). Conversely, supplementing the tRNA population with individual purified species may also lead to considerable frameshifting in vitro (Atkins et al., 1979b). Hence it may well be that the arrangement of the successive codons in the coding region of a natural mRNA as well as its corresponding overlapping triplets has evolved in order to minimize such frameshift errors in vivo. A study of the frequency of triplet usage in the non-reading frame of several highly expressed mRNAs showed that most codons that are systematically avoided in the normal reading frame (including the termination codons) are in fact quite often present in the +1 or +2 reading frames (Table IV) (see also Wain-Hobson et al., 1981). These codons correspond to rare tRNA and this may therefore limit the probability of a frameshift event from the normal reading frame to a nonsense frame.

TABLE IV

Codon usage in all three possible reading frames of the *tufA* gene of *E. coli*; correlation with anticodon abundance

The frequency of each codon used in the three overlapping frames of the *tufA* gene together with the anticodon sequence and corresponding relative abundance of the tRNAs are given. Codon frequencies are determined from the known sequence of the *tufA* gene (Yokota et al., 1980); zero in the column headings refers to the normal reading frame (123.123), +1 and +2 correspond to the frameshift triplets 231.231 and 312.312 respectively. Total number of codons analysed is 393. Relative amounts of tRNA and anticodon structure are also given. Data are normalized to 1.0 (underlined) for $tRNA^{leu}$ (anticodon CAG). These results are compiled from Ikemura (1981) and Post and Nomura (1980). Relative abundance for the minor $tRNA^{arg}$ corresponding to the codon AGA and presumably to AGG is taken from Caskey et al. (1968). The unique value (0.25) given by Ikemura (1981) for the $tRNA^{leus}$ which recognize the UUA and UUG codons is divided by two, because there are two minor tRNAs corresponding to each codon. Anticodons and their immediate 3′ (hyper) modified purine are known from sequence determination (see Gauss and Sprinzl, 1981), except for the five anticodons in brackets for which the identification comes from anticodon-anticodon binding experiments (Grosjean et al., 1978a; and unpublished results) or from gene sequencing (in lower-case letters) (An and Friesen, 1980). Those codons that are expected to exist but have not yet been unambiguously determined are indicated by symbol ?. V is uridine-5-oxyacetic acid; C^+ is N^4-acetyl cytidine; S is 5-methylaminomethyl-2-thiouridine; Q is 7-(4,5-*cis* dihydroxy-1-cyclopenten-3-ylaminomethyl)-7-deaza-guanosine; I is inosine, A *, G *, U * and Y are unidentified modified adenosine, guanosine, uridine and pyrimidine residues respectively. A^{si6} is N^6-(Δ^2-isopentenyl)2-methylthioadenosine; A^{t6} is *N*-9(β-D-ribofuranosyl)purin-6-ylcarbamoyl threonine; A^{6m} is N^6-methyl-adenosine and A^{2m} is 2-methyladenosine. Boxes indicate some of the triplets that are not used in the normal reading frame (except for a single occurrence of GGG) but that are clearly not avoided in frames +1 or +2. They are the terminator codons as well as codons for which we discuss a potential role in modulating the rate of translation (additional modulating codons may exist).

frame	Codon 0+1+2	Anticodon	Codon 0+1+2	Anticodon	Codon 0+1+2	Anticodon	Codon 0+1+2	Anticodon	
U	Phe	1 0 0 / 13 9 4 — 0.35 GAA.A^{si6}	Ser	7 4 7 / 3 10 5 — Major (GGA)	Tyr	2 1 6 / 8 2 13 — 0.5 QUA.A^{si6}	Cys	1 7 7 / 2 11 12 — ? GCA.A^{si6}	U
	Leu	0 4 1 - (012) $A^*AA.A^{si6}$ / 0 17 0 - (012)(CAA)		0 12 8 / 0 12 8 — 0.25 VGA.A^{si6}	ochre	**1 1 8** —	opal	**0 6 19** —	C
					amber	**0 6 0** —	Trp	1 18 13 - 0.3 CCA.A^{si6}	A
									G
C	Leu	1 2 4 / 1 11 2 — 0.3 GAG.G* / **0 8 6 - Minor** / 26 12 2 - 1.0 CAG.G*	Pro	0 2 7 / 0 3 4 — Minor / 1 7 5 / 19 8 8 — Major (VGG)	His	1 1 12 / 10 1 9 — 0.4 QUG.A^{2m}	Arg	21 3 17 / 2 6 9 — 0.9 ICG.A^{2m} / **0 10 17** / **0 8 14** -?	U
					Gln	0 3 13 - 0.3 SUG.A^{2m} / 8 2 0 - 0.4 CUG.A^{2m}			C
									A
									G
A	Ile	3 1 3 / 26 0 1 — 1.0 GAU.A^{t6} / **0 3 1 - 0.05 YAU.A^{t6}**	Thr	13 5 7 / 16 4 2 — 0.8 GGU.A^{t6} / 1 13 1 / 0 23 4 — Minor (tgt)	Asn	0 5 9 / 7 11 2 — 0.6 QUU.A^{t6}	Ser	0 5 6 / 0 6 3 — 0.25 GCU.A^{t6}	U
	Met	10 3 1 - 0.3 C*AU.A^{t6}			Lys	18 11 4 - 1.0 SUU.A^{t6} / 5 22 0 - ?	Arg	**0 7 13** (0.02) / **0 2 7** -?	C
									A
									G
G	Val	24 2 6 / 0 9 3 — 0.4 GAC.A / 10 12 3 / 3 18 2 — 1.05 VAC.A^{6m}	Ala	13 3 7 / 1 6 8 — Minor (GGC) / 5 2 6 / 8 14 3 — 1.04 VGC.A	Asp	4 0 12 / 21 0 6 — 0.8 QUC.A^{2m}	Gly	19 0 7 / 21 1 3 — 1.12 G CC.A / **0 0 12** - 0.15 U*CC.A / **1 1 7** - 0.10 CCC.A	U
					Glu	30 0 5 - 0.9 SUC.A^{2m} / 6 0 0 - ?			C
									A
									G
	U		C		A		G		

(2) The role of mRNA secondary structure

Not only the primary sequence but also the secondary and possibly the tertiary structure of an mRNA molecule may be important for an efficient in phase translation process. Unfortunately, little is known about the spatial organization of an mRNA molecule. However, from theoretical predictions as well as from biophysical measurements, it is clear that up to 70–80% of the bases in phage RNA and presumably in most RNAs are involved in base pairing (see Fitch, 1974; also Steitz, 1979; Fiers, 1979 and references therein). Such spatial configuration of the mRNA may limit its accessibility to the ribosomal subparticle and to the ini-

tiation factors so as to favour the initiation of protein synthesis at privileged sites of the mRNA (Steitz, 1979; Iserentant and Fiers, 1980). However, other parameters are also important for the initiation of protein synthesis in prokaryotes, such as the "Shine and Dalgarno sequence" (Shine and Dalgarno, 1975) and perhaps other characteristic sequences (Ganoza et al., 1978; Atkins, 1979; Sedlacek et al., 1978). A puzzling problem is how the ribosome manages to pass through regions of stable secondary structure of the mRNA during the elongation process. Presumably the translation machinery can unfold the helical regions of the mRNA. But experimental evidence indicates that the ribosomes may get stuck temporarily (i.e. there may be pauses) at certain regions of the mRNA. Whether such pauses result mainly from a "block" by the secondary structure (Chaney and Morris, 1979) or from an unbalanced codon-anticodon population, or whether they result from other mechanisms or constraints, is not clear at present. However, if secondary structure is important, one would expect that the constraints on the use of degenerate codons may also influence the efficiency of translation. This idea was advanced a long time ago (Adams et al., 1969; Min Jou et al., 1971), but up to now it has not been possible to prove or disprove it.

CONCLUSIONS

Not only does mRNA contain information for specifying a particular amino acid sequence, but its structure also determines the efficiency of translation. This is determined both by the frequency of initiation and—as discussed in this paper—by the proper choice of codons. The codon usage is markedly different in highly expressed genes compared with genes coding for rare proteins like repressors. Two different aspects of selective codon usage should be distinguished. Degenerate codons ending with U or C and recognized by the same tRNA are selected on the basis of an optimization of codon–anticodon interaction energy (neither too strong nor too weak). On the other hand, a number of modulating codons are recognized by minor tRNAs (or by a weakly interacting isoaccepting tRNA); these codons are clearly avoided in efficiently expressed genes.

Finally, it is important to stress that not only the expression level but also the *accuracy* of protein synthesis is an essential characteristic of the in-phase translation process. Certain successions of codons may be favoured because they limit the probability of missense or frameshift errors.

Some of the many factors which govern codon usage have been revealed for *E. coli*; possibly the same factors may also apply to other prokaryotes.

ACKNOWLEDGEMENTS

We thank Dr. G. Fayat, Dr. C. Sacerdot and Dr. S. Blanquet for having permitted us to use the sequence of IF3 before publication. H.G. benefits from grants from the Belgian Government "Actions Concertees" (to Professor H. Chantrenne) as well as from the "Fonds National de la Recherche Scientifique". W.F. is likewise supported by grants from the "Gekoncerteerde Onderzoeksakties" of the Belgian Ministry of Science, and from the "Fonds voor Geneeskundig Wetenschappelijk Onderzoek".

REFERENCES

Adams, J.M., Jeppesen, P.G.N., Sanger, F.V. and Barrell, B.G.: Nucleotide sequence from the coat protein cistron of R17 bacteriophage RNA. Nature 223 (1969) 1009–1014.

Alton, N.K. and Vapnek, D.: Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn*9*. Nature 282 (1979) 864–869.

An, G. and Friesen, J.D.: The nucleotide sequence of *TufB* and nearby tRNA structural genes of *E. coli*. Gene 12 (1980) 33–39.

Atkins, J.F.: Is UAA or UGA part of the recognition signal for ribosomal initiation? Nucl. Acids Res. 7 (1979) 1035–1041.

Atkins, J.F., Elseviers, D. and Gorini, L.: Low activity of β-galactosidase in frameshift mutants of *E. coli*. Proc. Natl. Acad. Sci. USA 69 (1972) 1192–1195.

Atkins, J.F., Steitz, J.A., Anderson, C.W. and Model, P.: Binding of mammalian ribosomes to MS2 phage RNA reveals an overlapping gene encoding a lysis function. Cell 18 (1979a) 247–256.

Atkins, J.F., Gesteland, R.F., Reid, B.R. and Anderson, C.W.: Normal tRNAs promote ribosomal frameshifting. Cell 18 (1979b) 1119–1131.

Beck, E. and Bremer, E.: Nucleotide sequence of the gene *ompA* coding the outer membrane protein II of *E. coli* K-12. Nucl. Acids Res. 8 (1980) 3011–3022.

Beremand, M.N. and Blumenthal, T.: Overlapping genes in RNA phage: a new protein implicated in lysis. Cell 18 (1979) 257–266.

Buchel, D.E., Groneborn, B. and Muller-Hill, B.: Sequence of the lactose permease gene. Nature 283 (1980) 541–545.

Buckingham, R.H. and Grosjean, H.: The accuracy of mRNA:tRNA recognition on the ribosome, in Galas, D.J. (Ed.) Accuracy in Molecular Biology. Dekker, New York, 1982, in press.

Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S. and Gross, C.A.: The nucleotide sequence of the cloned *rpoD* gene for RNA polymerase sigma subunit from *E. coli* K12. Nucl. Acids Res. 9 (1981) 2889–2903.

Caskey, C.T., Beaudet, A. and Nirenberg, M.: RNA codons and protein synthesis, 15. Dissimilar responses of mammalian and bacterial tRNA fractions to mRNA codon. J. Mol. Biol. 37 (1968) 99–118.

Chaney, W.G. and Morris, A.J.: Nonuniform size distribution of nascent peptides: the effect of mRNA structure upon the rate of translation. Arch. Biochim. Biophys. 194 (1979) 283–291.

Chavancy, G. and Garel, J.P.: Does quantitative tRNA adaptation to codon content in mRNA optimize the ribosomal translation efficiency? Proposal for a translation system model. Biochimie 63 (1981) 187–195.

Christie, G. and Platt, T.: Gene structure in the tryptophan operon of *E. coli* nucleotide sequence of *trpC*. J. Mol. Biol. 142 (1980) 519–530.

Cozzone, A.J.: Stringent control and protein synthesis in bacteria. Biochimie 62 (1980) 647–664.

Crawford, I.P., Nichols, B.P. and Yanofsky, C.: Nucleotide sequence of the *trpB* gene in *E. coli* and *S. typhimurium*. J. Mol. Biol. 142 (1980) 489–502.

Csordas-Toth, E., Boros, I. and Venetianer, P.: Nucleotide sequence of a secondary attachment site for bacteriophage λ on the *E. coli* chromosome. Nucl. Acids Res. 7 (1979) 1335–1341.

DiRienzo, J.M., Nakamata, K. and Inouye, M.: The outer membrane proteins of Gram-negative bacteria: biosynthesis, assembly and functions. Annu. Rev. Biochem. 47 (1978) 481–523.

Eisenberg, S.P., Yarus, M. and Soll, L.: The effect of an *E. coli* regulatory mutation on tRNA. J. Mol. Biol. 135 (1979) 111–126.

Farabaugh, P.J.: Sequence of the *lacI* gene. Nature 274 (1978) 766–769.

Fiers, W.: Structure and function of RNA bacteriophages, in Fraenkel-Conrat, H. and Wagner, R.R. (Eds.), Comprehensive Virology, Vol. 13, Plenum, New York, 1979, pp. 69–203.

Fiers, W. and Grosjean, H.: On codon usage. Nature 277 (1979) 328.

Fiers, W., Contreras, R., De Wachter, R., Haegeman, G., Merregaert, J., Min Jou, W. and Vandenberghe, A.: Recent progress in the sequence determination of bacteriophage MS2-RNA. Biochimie 53 (1971) 495–506.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D. and Merregaert, J.: Complete nucleotide sequence of bacteriophage MS2 RNA: Primary and secondary structure of the replicase gene. Nature 260 (1976) 500–507.

Fitch, W.M.: The large extent of putative secondary nucleic acid structure in random nucleotide sequences or amino acid derived mRNA. J. Mol. Evol. 3 (1974) 279–281.

Fitch, W.M.: Is there selection against wobble in codon-anticodon pairing? Science 194 (1976) 1173–1174.

Fox, T.D. and Weiss-Brummer, B.: Leaky +1 and −1 frameshift mutations at the same site in a yeast mitochondrial gene. Nature 288 (1980) 60–63.

Freier, S.M. and Tinoco, Jr., I.: The binding of complementary oligonucleotides to yeast initiator tRNA. Biochemistry 14 (1975) 3310–3311.

Gallant, J.A.: Stringent control in *E. coli*. Annu. Rev. Genet. 13 (1979) 393–415.

Gallant, J. and Foley, D.: On the causes and prevention of mistranslation, in Chambliss, G. Craven, G., Davies, J., Davis, K., Kahan, L. and Nomura, M. (Eds.), Ribosomes, Structure, Function and Genetics. Steenbock Symposium. University Park Press, Baltimore, 1979, pp. 615–638.

Ganoza, M.C., Fraser, A.R. and Neilson, T.: Nucleotides contiguous to AUG affect translational initiation. Biochemistry 17 (1978) 2769–2775.

Gauss, D.H. and Sprinzl, M.: Compilation of tRNA sequences. Nucl. Acids Res. 9 (1981) r1–r42.

Gouy, M. and Grantham, R.: Polypeptide elongation and tRNA cycling in *E. coli*: a dynamic approach. FEBS Lett. 115 (1980) 151–155.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R.: Codon catalog usage is a genome strategy modulated for gene expressivity. Nucl. Acids Res. 9 (1981) r43–r74.

Grosjean, H., de Henau, S. and Crothers, D.: On the physical basis for ambiguity in genetic coding interactions. Proc. Natl. Acad. Sci. USA 75 (1978a) 610–614.

Grosjean, H., Sankoff, D., Min Jou, W., Fiers, W. and Cedergren, R.J.: Bacteriophage MS2 RNA: a correlation between the stability of the codon: anticodon interaction and the choice of codewords. J. Mol. Evol. 12 (1978b) 113–119.

Gurevich, A.I., Igoshin, A.V. and Kolosov, M.N.: Structure of a central part of *E. coli* operon *rpoBC*: nucleotide sequence of the gene for β subunit of RNA polymerase. Biorg. Khim. (1980) 1580–1584 (in Russian).

Hedgepeth, J., Clement, J.M., Marchal, C., Perrin, D. and Hofnung, M.: DNA sequence encoding the NH2-terminal peptide involved in transport of λ receptor, an *E. coli* secretory protein. Proc. Natl. Acad. Sci. USA 77 (1980) 2621–2625.

Heffron, F., McCarthy, B.J., Ohtsubo, H. and Ohtsubo, E.: DNA sequence analysis of the transposon Tn*3*: three genes and three sites involved in transposition of Tn*3*. Cell 18 (1979) 1153–1163.

Horii, T., Ogawa, T. and Ogawa, H.: Organization of the *recA* gene of *E. coli*. Proc. Natl. Acad. Sci. USA 77 (1980) 313–317.

Ikemura, T.: Correlation between the abundance of *E. coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. 146 (1981) 1–21.

Iserentant, D. and Fiers, W.: Secondary structure of mRNA and efficiency of translation initiation. Gene 9 (1980) 1–12.

Kastelein, R.A., Remaut, E., Fiers, W. and Van Duin, J.: Lysis gene expression of RNA phage MS2 depends on a frameshift during translation of the overlapping coat protein gene. Nature 295 (1981) 35–41.

Lawther, R.P., Nichols, B., Zurawski, G. and Hatfield, G.W.: The nucleotide sequence preceding and including the beginning of *ilvE* gene of the *ilvGEDA* operon of *E. coli*. Nucl. Acids Res. 7 (1979) 2289–2301.

Min Jou, W., Haegeman, G. and Fiers, W.: Studies on the bacteriophage MS2: nucleotide fragments from the coat protein cistron. FEBS Lett. 13 (1971) 105–109.

Min Jou, W., Haegeman, M., Ysebaert, M. and Fiers, W.: Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature 237 (1972) 82–88.

Miyada, C.G., Horwitz, A.H., Cass, L.G., Timko, J. and Wilcox, G.: DNA sequence of the *araC* regulatory gene from *E. coli* B/r. Nucl. Acids Res. 8 (1980) 5267–5274.

Nakamura, K., Pirtle, R.M., Pirtle, I.L., Takeishi, K. and Inouye, M.: Messenger RNA of the lipoprotein of the *E. coli* outer membrane, II. The complete nucleotide sequence. J. Biol. Chem. 255 (1980) 210–216.

Nichols, B.P. and Yanofsky, C.: Nucleotide sequences of *trpA* of *Salmonella typhimurium* and *E. coli*: an evolutionary comparison. Proc. Natl. Acad. Sci. USA 76 (1979) 5244–5248.

Nichols, B.P., Miozzari, G.F., Van Cleemput, M., Bennett, G.N. and Yanofsky, C.: Nucleotide sequences of the *trpG(D)* regions of *E. coli, S. dysenteriae, S. typhimurium* and *S. marcescens*. J. Mol. Biol. 142 (1980) 503–517.

Nichols, B.P., Van Cleemput, M. and Yanofsky, C.: Nucleotide sequence of *E. coli trpE*: anthranilate synthetase component I contains no tryptophan residues. J. Mol. Biol. 146 (1981) 45–54.

Post, L.E. and Nomura, M.: Nucleotide sequence of the intercistronic region preceding the gene for RNA polymerase subunit α in *E. coli*. J. Biol. Chem. 254 (1979) 1064–1066.

Post, L.E. and Nomura, M.: DNA sequences from the str operon of *E. coli* J. Biol. Chem. 255 (1980) 4660–4666.

Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P.: Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit β in *E. coli*. Proc. Natl. Acad. Sci. USA 76 (1979) 1697–1701.

Post, L.E., Arfsten, A.E., Davis, G.R. and Nomura, M.: DNA sequence of the promoter region for the α ribosomal protein operon in *E. coli*. J. Biol. Chem. 255 (1980) 4653–4659.

Rose, J.K. and Yanofsky, C.: Interaction of the operator of the tryptophan operon with repressor. Proc. Natl. Acad. Sci. USA 71 (1974) 3134–3138.

Roth, R.R.: Frameshift suppression. Cell 24 (1981) 601–602.

Sancar, A., Stachelek, C., Konigsberg, W. and Rupp, D.W.: Sequences of the *recA* gene and protein. Proc. Natl. Acad. Sci. USA 77 (1980) 2611–2615.

Sanger, F., Air, G.M., Barrell, B.G., Brown, W.L., Coulson, A.R., Fiddes, J.C., Hutchison III, C.A., Slocombe, P.M. and Smith, M.: Nucleotide sequence of bacteriophage ϕX174 DNA. Nature 265 (1977) 687–695.

Sedlacek, J., Fabry, M., Rychlik, I., Volny, D. and Vitek, A.: Base pairing mRNA:rRNA: the arrangement of nucleotides in the message for coat protein of phage MS2. Acta Virol. 22 (1978) 353–361.

Shine, J. and Dalgarno, L.: Determinant of cistron specificity in bacterial ribosomes. Nature 254 (1975) 34–38.

Singleton, C.K., Roeder, W.D., Bogosian, G., Somerville, R.L. and Weith, H.L.: DNA sequence of the *E. coli trpR* gene and prediction of the amino acid sequence of Trp repressor. Nucl. Acids Res. 8 (1980) 1551–1559.

Smith, D.R. and Calvo, J.M.: Nucleotide sequence of *E. coli* gene coding for dihydrofolate reductase. Nucl. Acids Res. 8 (1980) 2255–2273.

Steitz, J.A.: Genetic signals and nucleotide sequences in mRNA, in Goldberger, R.F. (Ed.) Biological Regulation and Development, Vol. I. Plenum, New York, 1979, pp. 349–399.

Vacher, J., Buckingham, R.H., Houssier, U. and Grosjean, H.: Effect of ms^2i^6 modification in *E. coli* $tRNA^{Trp}$ on anticodon-anticodon binding: thermodynamic and kinetic evaluations. Arch. Int. Physiol. Biochem. 89 (1981) B204–B205.

Wain-Hobson, S., Nussinov, R., Brown, R.J. and Sussman, J.L.: Preferential codon usage in genes. Gene 13 (1981) 355–364.

Weiss, G.B.: Translational control of protein synthesis by tRNA unrelated to changes in tRNA concentration. J. Mol. Evol. 2 (1973) 199–204.

Weissenbach, J. and Grosjean, H.: Effect of t^6A in yeast $tRNA^{arg}_{Ifi}$ on codon–anticodon and anticodon–anticodon interactions: a thermodynamic and kinetic evaluation. Eur. J. Biochem. 116 (1981) 207–213.

Yokota, T., Sugisaki, H., Takanami, M. and Kaziro, Y.: The Nucleotide sequence of the cloned *tufA* gene of *E. coli*. Gene 12 (1980) 25–31.

Young, F.S. and Furano, A.V.: Regulation of the synthesis of *E. coli* elongation factor Tu. Cell 24 (1981) 695–706.

Communicated by A. Campbell.

Note added in proof

High levels of codon bias have now also been reported for highly expressed genes of *Saccharomyces cerevisiae* by Bennetzen and Hall (J. Biol. Chem. 257 (1982) 3026–3031). These authors point out that the preferred synonymous codons tend to be complementary to the anticodons of the major yeast isoacceptor tRNAs while codons with very strong or very weak pairing interaction are apparently avoided. Thus, in yeast as in *E. coli*, the rule of optimal codon–anticodon interaction energy seems to apply for highly expressed genes.