MICROBIOLOGICAL REVIEWS, June 1990, p. 198–210
0146-0749/90/020198-13$02.00/0

Vol. 54, No. 2

Codon Preferences in Free-Living Microorganisms

SIV G. E. ANDERSSON AND C. G. KURLAND*

Department of Molecular Biology, Uppsala University Biomedical Center, Box 590, Uppsala, Sweden S751 24

INTRODUCTION

The standard version of the genetic code includes 61 sense codons and three stop codons. Although almost all organisms have made the same codon assignments for each amino acid, the preferred use of individual codons varies greatly among taxonomic groups. For example, whereas the related bacteria *Escherichia coli* and *Salmonella typhimurium* have very similar codon preferences (51), the taxonomically unrelated *Bacillus subtilis* has a quite different preferential codon usage pattern (71, 89). In addition, a considerable heterogeneity exists within species; in this case, individual genes tend to favor characteristic codon distributions (5, 18, 35, 86, 89). In vertebrates, for example, codon choice depends mainly on the GC bias of the particular DNA segment harboring the gene (3, 6–8). In unicellular organisms there is a strong connection between protein expressivity and the degree of codon bias, which, in the extreme case, leads to the so-called major codon bias (5, 35, 37, 42, 86).

One of our principal aims here will be to account for the major codon bias of unicellular organisms such as *E. coli* and *Saccharomyces cerevisiae*. To do this, we discuss at some length the connections between codon usage and a variety of functional parameters such as the rates of translation, protein expression levels, and cellular growth rates.

Our starting point is the notion that the degeneracy of the genetic code may be used in ways that solve a number of different problems for the translation apparatus of a cell. In effect, the programming of amino acid sequences of proteins may be only the beginning of the list of functions for individual codons. For example, the degeneracy of the genetic code may be exploited to regulate gene expression or to modulate the performance of the translation system.

Furthermore, we do not assume that codon assignments for amino acids are permanent. Rather, we emphasize that codon assignments as well as codon preferences can replace one another systematically in response to the interplay of mutational pressure and the selective consequences of alternative assignments (18, 20, 53, 73, 74, 83–85; S. G. E. Andersson and C. G. Kurland, unpublished data). Here we wish to indicate how some of the selective pressures for the assignments of synonymous codons may be developed. To stress the dynamics of codon assignments and to illustrate the idea that the genetic code has a structure that can be related to its history, we will begin with a brief summary of current views of the origins of the genetic code.

* Corresponding author.

EVOLUTION OF THE CODE

No obvious clues to the composition of the primeval code are apparent in modern genomes. Nor do we know for sure which amino acids were first encoded by the original genetic systems. However, we may assume for heuristic purposes that the very earliest peptides were composed of the amino acids most abundant at that time. These in turn can be guessed from data about amino acid accumulation in atmospheres assumed to be prebiotic. One of the surprising observations here is that the products synthesized under such putative prebiotic conditions contain in high yield a relatively small number of amino acids (67). Likewise, studies of meteorites and their content of organic molecules reveal a small, related group of relevant compounds (67) (Table 1).

It is worth emphasizing that the most abundant amino acids in this group differ from the other amino acids abundant in modern proteins in that they are not synthesized from other amino acids (99). Thus, glycine, valine, aspartic acid, glutamic acid, and alanine are apparently both chemically and metabolically the simplest amino acids to accumulate. Accordingly, they may very well have been the most abundant amino acids in the primitive biosphere (67).

A basic requirement for the function of a primitive translation system lacking ribosomes is that the bond between the equivalents of the adapter and the messenger molecules should be strong enough not to come apart until the polypeptide chain is transferred to the amino acid attached to the next adaptor. It is therefore generally believed that the primitive RNA species involved in the early translation mechanism favored G+C-rich structures. According to Crick et al. (26), maximal binding stability would be promoted by a codon of the form RRY, assuming that the bases surrounding the anticodon are similar to those of present-day

TABLE 1. Relative abundance of amino acids in the Murchison meteorite and in an electric discharge synthesis[a]

Amino acid	Relative abundance in[b]:	
	Murchison meteorite	Electric discharge
Gly	++++	++++
Ala	++++	++++
Val	+++	++
Asp	+++	+++
Glu	+++	++
Pro	+++	+

[a] Data are from reference 67 (with permission). Molar ratio to glycine (=100): +, 0.05 to 0.5; ++, 0.5 to 5; +++, 5 to 50; ++++, >50.

TABLE 2. Biosynthetic relationship of amino acids[a]

Primary amino acid	Product amino acid(s)
Gly	
Ala	
Val	Leu
Asp	Lys, Asn, Thr
Glu	Pro, Arg, Gln
Ser	Trp, Cys
Phe	Tyr

[a] Data are from reference 99 (with permission).

tRNAs. Eigen and Winkler-Oswatitsch have favored a less extreme coding sequence consisting of repeating RNY triplets (31).

If the RNY code is combined with a preferred use of G's and C's, we might guess that the primitive code preferentially used GGC and GCC followed by GAC and GUC. It is suggestive that these codons now code for the amino acids most frequently synthesized in putative primitive atmospheres (Table 1). Apparently, the naïve arguments concerning the selective virtues of G · C pairs in primitive translation systems seem to be relevant. Furthermore, it has been observed that modern codon assignments tend to be enriched for RNN codons and in particular for GNN codons (87, 88, 92). Likewise, the amino acids derived directly from prebiotic synthesis have a higher frequency in today's proteins than those believed to have been derived from inventive biosynthesis (100). Finally, codon degeneracy is higher for the first group of amino acids than for the remainder (100).

These features of the code have been identified by some as the fossil remains of the primeval code (87, 88). In contrast, others have argued that this pattern of codon and amino acid preferences is a selected pattern (92, 100). Indeed, it must be said that even if the primitive genetic apparatus had preferred GNN codons, it is hard to believe that such a bias would have survived if it were not associated with some functional advantages to modern organisms; one such advantage will be discussed below.

CODON ASSIGNMENT

An intriguing feature of the genetic code is its virtual universality: with a few exceptions the same codon assignments are used in all organisms. Different proposals have been forwarded to explain this characteristic. At one extreme, Crick has suggested that modern codon assignments have been preserved as a frozen accident in which an original set of coincidental codon assignments have been conserved because the occurrence of amino acid replacements in proteins during a phase of shifting codon assignments would be lethal (25). In this hypothesis there is no obligatory relationship between the structure of the amino acids and their codon assignments. The other extreme is, of course, the conjecture that the modern code is universal because it is a functionally optimal code based on structural characteristics of the amino acids and their codon assignments (98). As discussed below, these are not exclusive accounts of the evolution of the genetic code.

A traditional way to explore the structural origins of the code is to try to find chemical correlations between anticodons or codons and their cognate amino acids. One such correlation would be to relate the hydophobicity of an amino acid to that of its cognate anticodon on the grounds that this might indicate a physical link between amino acids and their cognate adapters (59, 70). Unfortunately, previous searches for direct nucleic acid-amino acid interactions found little if any specificity in the weak complexes formed with mono-, di- and trinucleotides (70). In contrast, a very recent study has revealed sequence-specific binding of L-arginine to an intron within a self-splicing ribosomal precursor RNA from *Tetrahymena* spp. (101). Nevertheless, although amino acids, such as arginine, with a bulky, polar side chain may be able to interact stereospecifically with a nucleotide sequence, there is no evidence suggesting that the other amino acids in modern proteins can form site-specific complexes with short nucleotide sequences.

An alternative way to view the orderly development of the genetic code is to relate its evolution to that of amino acid metabolism (99). Thus, it can be argued that amino acids that were rare under prebiotic conditions, but could eventually be synthesized from other amino acids, would have entered the biosphere at a relatively advanced stage of its evolution. Such a biosynthetic elaboration could occur with the simultaneous reassignment of codons within the groups of biosynthetically related amino acids (99). The data, such as they are, support this scenario.

At present, roughly 80 different enzymes are required for amino acid biosynthesis. In several of these pathways, one amino acid is required as a precursor in the production of another amino acid. However, seven amino acids do not depend on other amino acids for their biosynthesis. These seven primary amino acids and their products are shown in Table 2. Five of these seven amino acids are those expected to have been present at a high concentration in the prebiotic soup (Table 1). Furthermore, the codons of these groups of precursor and product amino acids are internally related by single-base substitutions. Accordingly, it has been proposed that the addition of a new member to an amino acid group occurred concomitantly with the assignment of new codons (99).

Additional support for this view is found in the observation that some of the families of amino acids seem to be internally related via their respective activation systems. For example, a sequence comparison of $tRNA^{Phe}$ and $tRNA^{Tyr}$ reveals expected homologies (39). Likewise, sequence homology is found in comparisons of the Met and the Ile synthetases as well as between their tRNAs (15). Clearly, much more must be done to explore these homologies. Nevertheless, the available data suggest that at least part of the structure of the genetic code is a reflection of its coevolution with the metabolic pathways for the amino acids (99).

CODON REASSIGNMENT

It is generally believed that once a genetic system has reached a certain degree of sophistication, codon reassignments are prohibited because they would disturb the function of too many highly evolved proteins (25). Nevertheless, a number of deviant codon assignments have been found, in particular in mitochondrial genomes (33). Furthermore, the standard termination codons turn out to be less standard than might have been expected (33).

One account of the occasional variations of the code would be that these represent codon divergence that preceded the "freezing in" of the modern code (40, 65). However, sequence comparisons in *Tetrahymena* spp. for glutamine tRNAs that read the termination codons UAA as well as UAG, as well as those that read the standard glutamine codons, suggest that they diverged at a point that is well within the time since eucaryotes first emerged (52). Similarly, the observation that closely related mitochondria, such as, for example, the mold and the yeast mitochondria, have different assignment patterns (33) suggests that these have resulted from divergent evolution. Accordingly, rather than being a frozen fossil, the code seems to be evolving.

How do reassignments arise? Here, two mutually compatible views are relevant: one is that reassignments are selectively neutral events, and the other is that they are functionally selected events.

The neutral interpretation is emphasized in the codon capture hypothesis (74). Fundamental to this model is the idea that mutational bias of the replication system can drive the evolution of codon sequences (3, 57, 68, 73). According to this interpretation, not only are variations of genomic G+C content reflected in the usage frequencies of the codons, but also, in the extreme, they may lead to codon disappearance, particularly in small genomes (73, 74). The reappearance of these transient codons would then provide the opportunity for a second assignment, which may or may not be the original one (73, 74). Consistent with this interpretation is the correlation of a very high A+T content in the genome of *Mycoplasma capricolum* with its reassignment of the conventional termination signal UGA to a codon for tryptophan (69, 73).

In contrast, mitochondrial usage of UGA to code for tryptophan and of AUA to code for methionine is not restricted to A+T-rich genomes. In fact, both of these reassignments are found in genomes for which the A+T frequencies at the third codon position vary by more than 1 order of magnitude from one species to the other (Andersson and Kurland, unpublished). On the other hand, there is a strong under-representation of G's in mRNAs as well as in rRNAs, which may be correlated with the recruitment of UGA and AUA for tryptophan and methionine, respectively (Andersson and Kurland, unpublished). In other words, there is evidence that mutational bias influences the evolution of codon assignments, but the pattern of reassignments cannot be explained solely by this single factor.

Accordingly, it seems relevant to look for additional forces that might influence the patterns of codon reassignments. We note that a characteristic feature of many mitochondrial systems is their small genome size, which is correlated with the size of some of the corresponding gene products (4). For example, the vertebrate, insect, and echinoderm mitochondria have the smallest genomes as well as the smallest rRNAs and the smallest number of tRNAs. Significant here is the observation that it is in precisely these genomes that the reassignments of AGA, AGG, and AUA have developed (Andersson and Kurland, unpublished). It may therefore be useful to search for a connection between the codon reassignments and the pressure on genome size.

If tRNA species are suitably constructed, as, for example, by the removal of the modifications usually associated with the first nucleotide of the anticodon triplet, most of the four codon boxes defined by the variation of the third codon nucleotide can be translated by a single tRNA with the same designation as in the standard genetic code (4, 20, 73). In this way the number of different tRNA isoacceptor species is reduced to 24 in the smaller mitochondrial genomes. A further reduction is hampered by the standard codon assignments for the isoleucine, arginine, serine, and leucine families. However, the reassignments of the codons AGG and AGA from arginine to serine and that of the codon AUA from isoleucine to methionine eliminate the need for one of the two arginine tRNAs as well as for one of the two isoleucine tRNAs. This reduces the number of tRNA species to 22. Accordingly, we suggest that the reassignments of AGA, AGG, and AUA in mitochondria are selected to support a genome minimization strategy (20; Andersson and Kurland, unpublished).

How is the translational ambiguity of codon reassignment tolerated in the transition from one assignment to the other? First, we would expect codon reassignments to occur most often at the least frequently used codons because such a bias would minimize their disruptive effects on protein structure. Further, animal-mitochondrial genomes are rapidly evolving (16), which generates the conditions for rare codons to disappear and reappear in a reassigned form (74–76). And, mutatis mutandis, termination codons, which are normally used at low frequencies compared with sense codons, are highly favored for reassignment, but in this case there are additional virtues to be exploited.

When a termination codon is recruited by an aminoacyl-tRNA, the addition of the new amino acids would be expected to have a minimal effect on the performance of the modified protein unless the C-terminal sequence is critical for function (60). Furthermore, it is often observed that genes are punctuated with multiple termination codons. Here, too, the recruitment of one or another of the termination codons by an aminoacyl-tRNA would be expected to have only minimal effects on the functions of the resulting protein.

Finally, we would not expect that the reassignment of a codon would occur as a single mutational event, but, rather, we suggest that it would take place gradually in a series of mutational transformations of the relevant tRNA species. This means that the evolving translation system will not go instantly from one to another assignment of the codon. Instead, the transition will be accompanied by a period of ambiguous translation in which alternative interpretations of the reassigned codon will be expressed simultaneously. Recent work with bacterial mutants and antibiotics that raise the translation error frequencies has indicated that bacteria grow surprisingly well with dramatically enhanced error frequencies (32, 66). Indeed, there is mounting evidence that limited amino acid replacements most often have a minimal effect on the structure and performance of proteins (29, 58, 63).

In summary, we believe that there are acceptable, if vague, scenarios for the evolution of codon assignments even in modern organisms. It is against this background that we wish to view the evolution of extreme preferences for one or another group of synonymous codons. In particular, we favor the view that the evolution of codon reassignments in

mitochondria has been driven by the selection of a translation system that tends toward a minimal number of different tRNA species. Likewise, we will suggest that the evolution of gene-specific codon preferences is dependent on mechanisms to reduce the abundance of certain tRNA species under favorable growth conditions.

MAJOR CODON PREFERENCES

In general, the codon composition of unicellular organisms follows the base composition of the genome (68). A direct relationship was found between the G+C content of several species and the number of CNN anticodons (73). Accordingly, it has been suggested that for many genes the mutation bias of the DNA polymerases has been the main determinant for codon and anticodon composition (68, 73).

However, although the codon usage for most genes seems to reflect the average nucleotide composition of the genome as a whole, there is a subset of genes for which the codon choice is strongly biased toward a group of "major" codons (18, 86, 89). In the very highly biased group, containing, for example, genes for the ribosomal proteins, the elongation factors, and outer membrane proteins, there may be as much as a 100-fold variation in the usage frequency for preferred as opposed to avoided codons. The codon usage of such typically highly biased genes, as well as that of genes with a low codon bias, is shown in Tables 3 and 4. Furthermore, the major codon usage bias can be correlated with the expression level of the protein molecules (5, 35, 37, 42, 86). It is observed that the higher the protein production level, the higher the tendency to use only a subset of codons in the gene. In other words, for the highly expressed genes, codon usage seems not to be determined solely by the mutation rate, but seems to be under strong selection pressure (18, 83, 89).

The selective origin of the major codon preference is also apparent in comparisons of the mutation rates for different genes in a group of related enterobacteria (50, 85). Here, pairwise comparisons of homologous genes in, for example, *E. coli* and *S. typhimurium* reveal characteristic synonymous codon substitution rates: these are relatively low for genes with marked major codon preferences and more pronounced for genes with less highly biased codon preferences. In contrast, the nonsynonymous codon substitution rates are not strongly related to codon bias. Such results suggest that the synonymous codon preferences of the major proteins are constrained by selective forces operating most probably at the level of translation (83, 85).

A more direct indication that the constraints on synonymous codon preferences arise at the level of translation is found in the strong correlation observed between the major codons and the concentrations of their cognate tRNA species in the bacteria (34, 47–51). These data are illustrated in Fig. 1. In effect, a bacterium growing under normal laboratory conditions is synthesizing a protein population that is highly biased toward the products of a relatively small number of genes. Furthermore, the mRNA population coding for these proteins is constituted from a very biased subset of codons that are translated by a matching tRNA population dominated by a selected subset of isoacceptor species.

One extreme interpretation of the codon usage-tRNA isoacceptor correlation suggests that the codon composition of the mRNA pool has been "adjusted" to the isoacceptor distribution of the tRNA pool to provide a balanced flow of amino acids (46). The tacit assumption here is that although codon composition can be selected, tRNA concentrations are, for some undisclosed reason, immutable. In fact, tRNA isoacceptor concentrations are regulatable and respond, for example, to changes in growth conditions (V. Emilsson and C. G. Kurland, unpublished data; C. X. Fournier and A. D. McKay, personal communication). Thus, it seems that both tRNA isoacceptor levels and codon composition can evolve together (17), and the question remains; why do they do so in such a biased way?

TABLE 3. Codon usage in very highly biased and very lowly biased genes in *E. coli*[a]

Amino acid	Codon	Usage in: Highly biased genes	Usage in: Lowly biased genes	Amino acid	Codon	Usage in: Highly biased genes	Usage in: Lowly biased genes
Phe	UUU	45	339	Ser	UCU	150	138
	UUC	201	183		UCC	133	126
Leu	UUA	9	258		UCA	8	132
	UUG	10	240		UCG	14	160
	CUU	23	189	Pro	CCU	18	100
	CUC	32	170		CCC	3	97
	CUA	4	55		CCA	35	147
	CUG	585	658		CCG	238	281
Ile	AUU	92	433	Thr	ACU	177	156
	AUC	401	347		ACC	229	275
	AUA	1	95		ACA	13	118
Met	AUG	186	380		ACG	20	195
Val	GUU	379	275	Ala	GCU	347	228
	GUC	51	241		GCC	79	423
	GUA	196	166		GCA	238	332
	GUG	123	376		GCG	231	506
Tyr	UAU	59	260	Cys	UGU	19	68
	UAC	182	139		UGC	29	82
ter	UAA	23	24	ter	UGA	1	21
	UAG	0	2	Trp	UGG	50	185
His	CAU	25	188	Arg	CGU	325	284
	CAC	111	125		CGC	109	360
Gln	CAA	33	293		CGA	0	86
	CAG	274	461		CGG	1	97
Asn	AAU	18	327	Ser	AGU	6	159
	AAC	308	283		AGC	75	270
Lys	AAA	470	441	Arg	AGA	0	41
	AAG	125	167		AGG	0	26
Asp	GAU	166	441	Gly	GGU	439	365
	GAC	320	205		GGC	300	341
Glu	GAA	451	590		GGA	4	150
	GAG	115	308		GGG	8	171

[a] Data are from reference 18 (with permission).

OPTIMAL CODONS

It seems natural to identify the major codon preference with a subset of particularly "good" codons. One theory to account for the superiority of these codons focuses on the notion that the stability of the codon-anticodon interaction must be optimized (42, 43). According to this view, a tRNA with the capacity to interact with two or more isocodons will be optimally matched when the G+C content of the codon-anticodon interaction is intermediate between the extremes of weak A · U and strong G · C pairings.

This physical model is contradicted by several observations. First, direct physical measurements of the stabilities of interactions between the complementary anticodons of suitable tRNA pairs reveal that the differences between G+C-rich and A+U-rich pairings are reduced by the loop

TABLE 4. Codon usage in very highly biased and very lowly biased genes in *S. cerevisiae*[a]

Amino acid	Codon	Usage in: Highly biased genes	Usage in: Lowly biased genes	Amino acid	Codon	Usage in: Highly biased genes	Usage in: Lowly biased genes
Phe	UUU	6	126	Ser	UCU	91	141
	UUC	64	64		UCC	64	63
Leu	UUA	15	130		UCA	5	83
	UUG	152	142		UCG	0	47
	CUU	2	64	Pro	CCU	3	56
	CUC	0	46		CCC	1	33
	CUA	6	74		CCA	78	55
	CUG	0	59		CCG	0	25
Ile	AUU	61	171	Thr	ACU	51	105
	AUC	61	75		ACC	63	55
	AUA	0	116		ACA	0	100
Met	AUG	39	103		ACG	0	47
Val	GUU	103	102	Ala	GCU	184	80
	GUC	87	70		GCC	49	67
	GUA	0	75		GCA	1	87
	GUG	1	66		GCG	0	43
Tyr	UAU	4	98	Cys	UGU	18	47
	UAC	55	84		UGC	1	21
ter	UAA	8	4	ter	UGA	0	6
	UAG	0	5	Trp	UGG	22	55
His	CAU	9	53	Arg	CGU	12	30
	CAC	39	34		CGC	0	16
Gln	CAA	63	165		CGA	0	30
	CAG	0	90		CGG	0	9
Asn	AAU	3	255	Ser	AGU	4	74
	AAC	88	155		AGC	0	62
Lys	AAA	33	251	Arg	AGA	99	108
	AAG	191	169		AGG	0	64
Asp	GAU	32	204	Gly	GGU	179	123
	GAC	91	107		GGC	4	61
Glu	GAA	131	218		GGA	0	54
	GAG	2	106		GGG	0	36

[a] Data are from reference 18 (with permission).

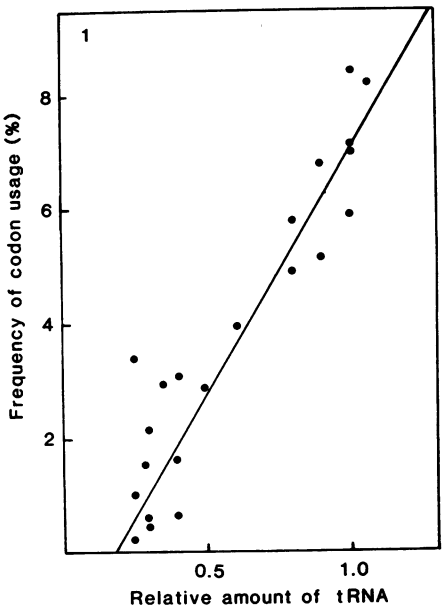

FIG. 1. Relationship between the frequency of codon usage in very highly biased genes and the content of isoaccepting tRNAs. The codon frequencies have been taken from reference 18, and the amount of tRNAs has been taken from reference 50 (with permission).

constraint of the anticodon (41). Second, the predicted difference in translational efficiency between G+C-rich and A+U-rich synthetic mRNA species is not observed in vitro (2). Third, the postulated bias toward codon-anticodon pairs of intermediate stability is absent in bacteria as different as *E. coli* (18), *B. subtilis* (89), *M. capricolum* (69), and *Micrococcus luteus* (72). Finally, different organisms use different subsets of major codons (36–38), which argues strongly against there being any intrinsic characteristic of the codons that can be the basis of their preferred character. In fact, virtually every single codon is used as a major codon by some system and as a minor codon by another.

It seems reasonable to assume that codon composition in the highly expressed genes is under the same mutational pressure at the DNA level as it is in the weakly expressed genes. In fact, the G+C content of the third codon position of ribosomal protein genes varies by as much as a factor of 10 in different bacteria. However, in each individual case that has been studied, this variation is biased in the same direction as the G+C content of the whole genome (68). Furthermore, highly expressed and poorly expressed genes in *E. coli* have very similar G+C contents in the third codon position, even though they have different codon preferences. This strongly suggests that a pronounced mutational pressure has been exerted uniformly on all of the genome. In fact, in some of these species, such as *M. capricolum* and *M. luteus*, it appears that mutational biases completely dominate codon usage.

None of this is meant to imply that there are no translational differences among codons. Rather, we suggest that an optimization of translational efficiency is possible for any arbitrarily chosen codon, for example by adjusting tRNA concentrations. Therefore, there is no reason to suppose that the properties of the codon-anticodon interaction provide the sole grounds for the selection of codon preferences.

REGULATORY CODONS

A favorite class of theoretical codon preference strategies postulates a role in the regulation of gene expression for a subset of codons. One particularly persistent rumor concerns the regulatory functions of the most infrequently used codons, the so-called rare codons. This conjecture has at least two forms. One is that rare codons modulate the expression levels of proteins present in low copy numbers (42, 44, 93). The other is that regulatory proteins themselves are encoded preferentially by rare codons to keep their expression levels low (55) and, similarly, that rare codons have been selected in signal sequences to reduce the rate of translations (21). Both forms of this conjecture are insupportable.

Thus, a recent compilation of codon usage data shows that regulatory genes and signal peptides do not have a significantly higher frequency of rare codons that a great number of other genes, expressed at moderate to low levels (84). In fact, even in these latter genes there is a slight tendency to avoid rare codons, although this bias is much less pronounced than in highly expressed genes. Most important, no gene has been sequenced yet that shows a preference for rare codons. Accordingly, although rare codons are avoided by major proteins, they are not preferred by minor ones.

Similarly, an inverse relationship between the occurrence of rare codons and the yield of protein is assumed in these

models. This expectation is based on the notion that poorly translated codons will lower the expression levels. However, as we shall see below, there is no necessary relationship between expression level and translation rate. In fact, high expression levels are routinely achieved for genes with a relatively high content of rare codons. For example, even though the chloramphenicol acetyltransferase gene has a codon usage very different from that of an optimal *E. coli* gene, it can be expressed to high levels when cloned under a strong promoter (82). Similarly, bacteriophage lambda can be highly expressed when cloned under a strong promoter, even though its codon usage is similar to that of the weakly expressed host genes (46). The same phenomenon has been observed in *S. cerevisiae*, in which hepatitis B virus core antigen was expressed to a level of 40% of soluble proteins even though it has a very low index of codon bias (54). In this case, it was shown that the high expression level was very dependent on 5'- and 3'-flanking sequences derived from yeast sequences.

Finally, if rare codons were selected for encoding regulatory proteins, they would be expected to have relatively low mutation rates in the corresponding genes. However, two such regulatory genes, *dnaG* and *araC*, accumulate synonymous substitutions at a rate matching that of genes with little codon bias, which contradicts the expectation of the regulatory model (85). On the other hand, rare codons may have an effect on attenuation by being clustered in the leader sequences (97).

We can turn the problem on its head by anticipating a different sort of codon strategy, in which a subset of genes are very highly expressed under certain growth states. We find that these genes are selected with a strong preference for a subset of codons (the major codon bias). Next, we consider the fate of the codons that are avoided in this subset of genes, assuming that there are no other codon preferences at work. First, it seems clear that all codons must be translatable in the system so that random mutations do not lead to lethal events with significant probability. However, this constraint need not be an extreme one. Instead, we would expect that the selection pressure to retain rare codons is slight in weakly expressed genes. Supporting this view is the finding of an inverse relationship between codon usage bias and the rate of synonymous nucleotide substitutions (85). Thus, it seems likely that rare codons in regulatory and weakly expressed genes result from the absence of strong negative selection rather than from the presence of strong positive selection.

EXPRESSION LEVEL IS NOT TRANSLATION RATE

All of this is not to say that certain codons do not influence translation rates or that it is, in principle, not possible to regulate protein expression levels by the judicious use of certain codons. However, even if an mRNA is constituted by a large number of "slow" codons, it will not necessarily be expressed at a low level, even if its translation rate is significantly depressed by its codon complement. The point is that the overall rate of which an mRNA is translated can directly influence the expression level of its polypeptide product if and only if the mRNA species can capture a major fraction of the ribosomes in the cell. To understand this constraint, we shall consider the relation of translation rate and expression levels in a little more detail.

In the steady state that corresponds to exponential cell growth, the expression level for a particular gene will be determined by the rate of synthesis of its mRNA and the number of ribosomes that complete the translation of each of these mRNA species. Under normal conditions, even the most highly expressed genes produce an mRNA pool that is only a minor fraction of the total mRNA population. In other words, if the mRNA pool of a cell consisted of only one mRNA species, an increase in the speed of its translation would lead (all other things being equal) to a higher expression level for the corresponding gene product, because the number of ribosomes that translate the mRNA per unit time is increased. However, when we consider the influence of an increase in the speed of translation of an mRNA that represents a very small fraction of the total mRNA pool, this effect is reduced to a corresponding degree. The reason is that a ribosome which has completed translation of an mRNA that is a minor fraction of the total mRNA pool will, with overwhelmingly high probability, be captured afterwards by a different mRNA species. This means that the average number of ribosomes that translate a particular mRNA per unit time will not be influenced by a variation of its rate of translation. Accordingly, the rate of translation of mRNA species will normally not influence the expression level of the corresponding protein.

On the other hand, it is in principle possible, through a judicious selection of codons at the beginning of the mRNA, to influence the queuing of ribosomes on the mRNA; this could lead to a modulation of the expression level for the corresponding proteins (19, 61). In this case we imagine that some codons are translated quickly and some are translated slowly. Accumulating the slower codons at the beginning of the gene sequence would tend to lower the ribosomal loading frequency, and accumulating the fast codons at the beginning would tend to raise the loading frequency. The finding that rare, presumably slow codons are accumulated preferentially at the beginning of a few genes that are normally expressed at a low level is consistent with such a queuing model for the regulation of gene expression by a biased codon sequence (61). However, a general review of a larger number of genes in *E. coli* as well as in the yeast *S. cerevisiae* has revealed that in general codon bias is less extreme at the beginning of genes than further on in the sequences; indeed, there is a remarkable similarity in the low codon bias of the initial sequences in weakly, moderately, and highly expressed genes (19). It therefore seems highly unlikely that these organisms regulate gene expression by codon-modulated ribosomal queuing.

CODON-SPECIFIC RATES

We have completed some of the preliminaries, and now we shall focus our attention on some relevant molecular details of the translation machinery as well as their relationship to cellular growth rates. The reason for this combination of concerns is that organisms such as enterobacteria and certain yeasts are the ones that are characterized by a well-developed major codon preference. That they also represent organisms for which competitive cellular growth rates provide a significant selective advantage is an important clue to understanding the major codon strategy.

From the very first observations of a direct relationship between codon bias and relative concentration of cognate tRNA species, it was tempting to guess that the rates of translation at particular codons were proportional to the amounts of cognate tRNA present in cells (22, 34). In contrast, microbiologists were taught by the Copenhagen school that biosynthetic efficiency always required ribosomes to work at maximum rates (62). Since maximum rates

are obtained only in the presence of excess substrate, it seemed, from this point of view, that translation rates could not be limited by tRNA availability. However, it might be that intrinsic differences exist for the maximum rates of translation at different codons (42, 43, 50). Accordingly, the questions that naturally arise are as follows. Do translation rates vary from codon to codon? Do ribosomes normally translate at their maximum rates for different codons? How different are these rates? To what extent are codon-dependent rate differences due to limiting tRNA concentrations?

The first clear indication that codons are translated at nonuniform rates came from studies of so-called pause sites for proteins that are destined for export through bacterial membranes (81, 94). Thus, for some of these proteins the translation process is markedly uneven, and incomplete polypeptide intermediates accumulate at sites where the ribosomes are retarded in their transit of the mRNA. It could be shown that such pause sites are correlated with short strings of seldom-used or rare codons (94, 95). It was concluded that these particular codons are translated slowly because their tRNA species are in short supply and that the selection of aminoacyl-tRNA to match these codons is rate limiting for peptide bond formation (48, 94). In this case the alternative interpretation that ribosomes translate in vivo at rates independent of the tRNA concentration but dependent on the identity of the codons was not favored.

A number of other studies with modified gene sequences have shown that the introduction of strings of rare codons can lower the expression level of the corresponding protein products (82). Similarly, codon substitutions in the leader peptide of *pyrE* affected the frequency of transcription past the attenuator (11–13). The interpretation of these results was that rare codons regulate gene expression level by lowering translation rates. However, in none of these studies has the rate of translation been measured. Since we have seen there that there is no automatic connection between expression level and translation rate, the meaning of the results of these experiments is unclear. For example, none of these studies could distinguish the effects of the codon insertions on translation from those on transcription or mRNA half-life. In fact, replacing an increasing number of major codons with synonymous minor ones in the phosphoglycerate kinase gene expressed in *S. cerevisiae* results in a drastic decrease in the steady-state level of the corresponding mRNA (45).

However, it has been recognized that strings of rare codons can in principle act synergistically to reduce their own translation rates (95). In this case the rare codons must be serviced by comparably rare tRNA species; the concentrations of these tRNA species must be rate limiting for translation, and the steady-state concentration of the rare codon strings being translated must be greater than those of their cognate tRNA species. If these unnatural conditions are met, as, for example, with a high-copy-number plasmid construction, the sequestering of the tRNA species at the codons in the string will amplify the translational delay that individual rare codons manifest. In addition to the sequestering effects, clusters of repeated rare codons may produce translational movement errors; indeed, a frameshift frequency of 50% was recently observed following the insertion of two neighboring AGG codons (91). Either or both of these effects may have played a role in the earlier experiments concerning the influence of short strings of rare codons on expression levels.

Genuine elongation rate measurements have been carried out for proteins encoded by mRNA species with different codon usage biases. These data reveal a clear positive correlation between the elongation rate and the prevalence of major codons (78). Another sort of experiment with genes containing inserted codon strings has provided additional evidence that translation rates vary from codon to codon (90). Since these data, as incomplete as they are, provide the most direct measurements to date of codon-specific translation rates, we shall inspect them more closely.

First, the inserts to be tested were versions of a ca. 20-codon sequence excised from a ribosomal gene; the virtue of this particular sequence is that when it is inserted into a gene in one reading frame it functions as a string rich in commonly used codons, whereas in another reading frame it provides a string rich in infrequently used codons. By measuring the delay in the transit time of the bacterial ribosomes caused by the inserts, it could be shown that common codons are translated at an average rate of 12 amino acids per s, whereas the uncommon ones are translated with an average rate of 2 amino acids per s (90).

The sixfold difference in translation rate observed for commonly used and seldom-used codons in these experiments is almost certainly an underestimate of the rate extremes for different codons, because it measures only averages for strings of codons. Thus, the biggest shortcoming of this sort of measurement is that it cannot yet be applied to individual codons. Nevertheless, these data show that translation rates at different codons can vary quite significantly. We next attempt to estimate the extent to which this variation is due to differences in tRNA concentrations or to the codon dependence of maximal translation rates.

In this context it is worth noting more recent experiments by Sörensen and Pedersen (Abstr. 13th Int. Transfer RNA Meet. 1989, abstr. no. mo-am-13). They have suggested that the same tRNA species may translate two isocodons at very different rates. They have measured the translation rate of strings of GAA, a major codon for Glu, compared with that of GAG, an uncommon codon for Glu. Both of these are thought to be translated by $tRNA_2^{Glu}$, but the GAA string is translated at least three times faster in vivo than is the GAG string. Such observations provide the most convincing evidence that translation rates can be codon specific. Nevertheless, for our purposes it is the tRNA concentration dependence of the translation rate that is most relevant.

HUNGRY RIBOSOMES

The suggestion that ribosomes function optimally when they are driven at their maximum rates is based on the notion that the mass investment in the ribosome is much greater than that of any other component of the translation system (62). Indeed, if the mass investment in all of the other components of the system is negligible, the rate of translation per ribosome should be the only factor contributing to the rate optimization of the translation system. However, if the mass investment into the aminoacyl-tRNA ternary complexes with elongation factor Tu (TF-Tu) and GTP is assumed not to be negligible, the rate optimization of the translation system is more complex.

If the masses of ternary complex are included in the optimization, the Maaløe maximization of the ribosome rate is found to be one theoretical limit of the optimization (30). It corresponds to the extreme of the highest conceivable growth rates in which the organism is doing little else than synthesizing proteins. At the other extreme of the lowest conceivable growth rates, the organism is doing very little

protein synthesis and is engaged mainly in building up amino acids, nucleotides, and other building blocks from a very simple medium. Here, the optimal mass investment is an equipartitioning of ternary complex and ribosome mass, and the kinetics of translation are limited by the concentrations of ternary complexes and ribosomes (30).

In reality, laboratory culture conditions support bacterial growth in states that are intermediate between these two theoretical extremes. The culture conditions supporting the fastest bacterial growth may begin to approach the extreme of translation rates limited by the maximum turnover rates of ribosomes. However, the expectation is that normal laboratory media support growth states in which the rates of translation at most codons are responsive to variations in the intracellular concentrations of cognate aminoacyl-tRNA ternary complexes (30). Furthermore, it has been observed that the presence in vitro of enormous excesses of a noncognate aminoacyl-tRNA ternary complex that is inclined to make errors at a particular codon has a vanishingly small effect on the translation rate at that codon by the cognate species (9). Accordingly, everything points to the steady-state concentration of ternary complex as the parameter that uniquely determines the translation rate at its cognate codon for a particular bacterial translation system.

Support for this view has come from experiments with ribosome mutants. Thus, a series of mutant bacteria with altered ribosomes were identified that could be ordered with respect to their rather different growth rates, and it was found that they were characterized by a proportional ordering with respect to their translational elongation rates in vivo (1). This correlation suggests that the rate of growth in these mutants is limited by the translational elongation rates. When the ribosomes from these mutants were isolated and studied in vitro, their maximum turnover rates were not very different from those of wild-type ribosomes. This suggests that the lower rates of translation by the mutant ribosomes is not due to lower maximum turnover rates, which is equivalent to saying that they are not saturated by tRNA-ternary complex in vivo. Instead, it was observed that the K_m value which describes the concentration of ternary complex necessary to achieve half of the maximum ribosomal turnover rate during translation was systematically altered in all of the mutants. In particular, the rate of translation in vivo was inversely proportional to the K_ms of the mutant ribosomes measured in vitro (1). This means that the affinity of the cognate ternary complexes for the different mutant ribosomes is positively correlated with the rate of translation in vivo. This is kinetically equivalent to saying that the ribosomes are not saturated by ternary complex in vivo and, accordingly, that the translation rates of these bacteria are dependent on the ternary complex concentrations in vivo.

The validity of this conclusion depends on the reliability of two sorts of extrapolations. One concerns the reliability of conclusions based on combinations of measurements made in vitro and in vivo. The other concerns the extrapolation from results obtained in vitro with one codon to all the other codons being translated in vivo. These are recognizable dilemmas to the biochemist, and their resolution depends on the outcome of further tests carried out in vivo. Here, the situation is encouraging.

An assay has been developed to compare the aminoacyl-tRNA selection rates at different codons in vivo. It takes advantage of an unusual sequence in the mRNA for the RF-2 protein of *E. coli* (23, 24). When this mRNA is translated in one reading frame, the polypeptide is aborted at a termination codon in the middle of the mRNA. Alternatively, the ribosome may shift reading frames and, in the new phase, complete the translation of the functional form of RF-2, which occurs at extraordinarily high frequencies in this particular sequence. In effect, there is a reading frame branch point in this mRNA which was exploited by Curran and Yarus (27, 28) to study the codon dependence of the initial selection kinetics in vivo. Thus, they have shown that when a sense codon replaces the nonsense codon in this sequence, its recognition by a tRNA species is competitive with the events leading to the frameshift. Accordingly, they could use this branch point to compare the initial kinetics for tRNA selection at different codons by measuring the relative frequencies of codon translation and frameshifting in a series of suitable mRNA constructions.

The results obtained in this assay for 29 different codons can be summarized as follows. First, the initial rates of codon recognition vary quite significantly, and within this group of codons the extremes vary as much as 25-fold. Of special interest is a reasonably good correlation between the more rapidly recognized codons and the commonly used codons. However, for the minor codons the situation is more complex. The frequencies for some minor codons can be correlated with the initial kinetics of the selection, but for the majority there is no correlation. Furthermore, the relative codon recognition rates for isocodons recognized by the same tRNA species were compared. These initial kinetics tend to be very similar; in no case did the relative frequencies for two isocodons differ by more than a factor of 2.

Undeniably, the most disconcerting aspect of the experiment is that it relates information about two undefined events which are nested within the complete peptide elongation cycle. One of these events is part of the initial discrimination step for aminoacyl-tRNA acquisition. The other is the event that initiates the unusual high-frequency frameshift event characteristic of this sequence. This is almost certainly not a normal part of the peptide elongation cycle, because it requires the participation of a Shine-Dalgarno sequence; accordingly, it is not clear what the frameshift event is reporting. In summary, we are not getting an unambiguous signal that reports information about the whole peptide elongation cycle from these experiments. Nevertheless, the experiment is state of the art.

When the ribosome is translating at its maximum rate, the EF-Tu-dependent steps leading to peptide bond formation and the translocation steps mediated by EF-G take up roughly equal parts of the peptide cycle (10). However, when the cognate ternary complex is below saturating concentrations, the time required to make the peptide bond will lengthen because the acquisition time for the tRNA has lengthened. As a consequence, the measurements made by Curran and Yarus (28) for rare codons translated by tRNA species that are equally rare more accurately reflect the time required to complete a peptide bond than do the measurements made on the major codons translated by tRNA species that are at the highest concentrations. Accordingly, we can be reasonably confident from the data on the slowest codons that there really is a large variation of codon-dependent translation times. This supports the conclusions drawn by Pedersen and colleagues (78, 90) concerning the codon dependence of elongation. Likewise, for the subset of relatively slow codons whose translation times are correlated with the relative abundance of their cognate tRNAs, it is not unlikely that the translation times are determined primarily by the availability of a matching ternary complex.

In summary, all the available data show that there is a great range of translation times for different codons. The

analysis of the ribosome mutants with altered kinetic characteristics provides strong support of the interpretation that translation rates for most codons are limited by the availability of the cognate tRNA species. In addition, there seem to be intrinsic kinetic differences in the rates of recognition and translation for some of these codons. However, the fact that two isocodons have different translation rates does not mean that they are not rate limited as well by the availability of their shared tRNA species. For another group of codons, the translation rate clearly seems to be determined by the degree to which individual codons are starved for their cognate ternary complexes. In particular, the major codons belong to the group that is read by the most abundant tRNA species. According to Pedersen and colleagues (78, 90), these are translated at the higher rates, and according to Curran and Yarus (28), they are recognized at the higher rates. Our next concern will be to relate the speed of translation to a selective parameter that will account for the major codon preference.

GROWTH MAXIMIZATION STRATEGIES

We noted above that if there were only one mRNA species to be translated by a fixed number of ribosomes, the number of proteins produced per unit time in the steady state would be proportional to the average rate of translation per codon. This is so because the availability of ribosomes to start a new polypeptide is influenced by the speed with which they complete the transit of the mRNA. Therefore, in this case, the use of fast rather than slow codons would make a difference in the translational efficiency. Similarly, a heterogeneous collection of mRNA species that used the same subset of codons would be translated faster or slower depending on whether the codon bias is for fast or slow codons. We wish to suggest that this sort of coupling between codon bias and translation rate is the selective virtue of the major codon preference in organisms such as *E. coli*. To see how this would work, we must introduce a connection between translational efficiency and growth rates.

The biosynthetic machinery of a cell reproduces itself as well as the other working components of the cell. If the flow of amino acids into proteins under particular growth conditions is limited, the time required to produce the proteins of the cell will be minimized if the rate of translation is maximized and the mass of the translation machinery is minimized. In other words, the growth rate will be influenced not simply by the rate of translation but by this rate normalized to the mass of the translational machinery (30).

In addition, the extent to which translational efficiency influences growth rate will depend on the fraction of the total biomass that is invested in translation (30). This means that the significance of translational efficiency for the growth rate depends on the growth conditions. At high growth rates in rich media, translation is the dominant cellular function of a bacterium, and, accordingly, translation efficiency should be a definitive parameter. At very low growth rates in poor media, the translation machinery is a small fraction of the cell mass, and the translational efficiency should have a correspondingly small influence on the growth rates. This view of the variable influence of translational efficiency on growth rate has been directly verified with a series of ribosome mutants. It has been shown that the growth rate impairment due to defective ribosomes is maximal at the highest growth rates and decreases systematically as the quality of the growth medium is lowered (66).

A related parameter that is growth rate dependent is the heterogeneity of the protein population (62). Under poor growth conditions a broad spectrum of enzymes as well as core components such as translational, transcriptional, and membrane proteins are being produced. In contrast, at the highest growth rates the same core components completely dominate the protein population. This means that the mRNA population will become progressively less heterogeneous as the growth rates increase, and at the highest growth rates the species that code for core components will represent most of the mRNA species. If we suppose that only a subset of codons are used preferentially to code for the core proteins, a situation is created that could markedly influence the efficiency of the translation system at the highest growth rates (57).

Maximization of the rates of translation would require increasing the aminoacyl-tRNA ternary complex concentrations sufficiently to drive the ribosomes at near-maximum rates. However, if this mass increase were uniform for all ternary complexes, it would tend to lower significantly the efficiency of translation. On the other hand, if the core proteins were encoded preferentially by a subset of codons, the mass investment in ternary complexes corresponding to these codons could be compensated, at least in part, by decreasing the relative concentration of the ternary complexes corresponding to the codons that are avoided in the core protein genes (57).

The major proteins of bacteria grown under normal laboratory conditions are, in fact, those that we have identified as the core proteins, and their relative amounts decrease substantially in very slowly growing cells. If the major codon preference is, as we suggest, a strategy to maximize growth rates in relatively rich media, the high-concentration tRNA species that are matched with the major codons in rapidly growing (normal) cultures should diminish significantly at the very lowest growth rates. In other words, the growth optimization strategy requires that the steady-state concentrations of individual tRNA species are regulated in a growth rate-dependent manner.

The data concerning the isoacceptor levels at different growth rates are far from complete, but the fragmentary data available at this writing are encouraging. First, the relative amounts of the different tRNA species under different growth conditions are not constant, but vary in systematic ways (Emilsson and Kurland, unpublished; Fournier and McKay, personal communication). For example, analysis of the leucine family shows that the concentration of the major Leu isoacceptor progressively increases at higher growth rates. Likewise, the level of the one minor Leu species that also can translate the major Leu codon ($tRNA_3^{Leu}$; CUA, CUG) increases almost threefold, whereas the remaining three decrease by factors of 2 to 3 as the growth rate increases (Emilsson and Kurland, unpublished). In the same series of measurements it was found that the levels of all three members of the methionine acceptor family increase when the growth rates are increased. Since these read AUG, they are by definition translators of a major codon, and their increase with growth rate is the predicted behavior.

In summary, there are reasonable grounds to believe that the major codon preference is a strategy to adapt bacteria and yeasts to rapid growth. Although it is not a strategy to regulate protein expression levels, there is indeed a very important regulatory problem nested in the major codon strategy. This concerns the unknown mechanism that regulates in a growth-rate-dependent way the expression level for each individual tRNA isoacceptor species.

DIVERSE REMEDIES

Roughly speaking, there are three aspects of protein biosynthesis that can go wrong: (i) individual amino acid missense substitutions can occur; (ii) processivity can be interrupted by a reading frame error, a drop-off event, or an abortive termination event; and (iii) protein folding can go awry. There are suggestions that specific codon preference strategies are directed at minimizing these sorts of errors.

It should be said that although, as we have claimed above, cells are more tolerant of translational ambiguity than was initially expected, this does not mean that translational errors are not deleterious. Rather, we mean that tolerable mistakes are still deleterious in the sense that they represent a wastage of biosynthetic potential. Accordingly, strategies to minimize such wastage will be advantageous to cells.

Unfortunately, so few data are available that virtually nothing can be said about the codon dependence of the missense errors in vivo (77). However, it is clear enough that whenever a minor codon can be misread by a major tRNA species, there is the danger that unacceptably high missense error rates will be the result. This sort of consideration has led to the suggestion that the choice of the members of the major codon preference is determined by a preference for major isoacceptor species that would minimize the missense errors caused by unequal cognate and noncognate tRNA concentrations (64). The search for such an error minimization strategy is yet another good reason to explore the details of the error rates with different tRNA species at different codons. Such a program must await more sensitive and less demanding procedures than are now available for the determination of missense substitutions in vivo (14).

There is general agreement that primitive analogs of mRNA must have contained some kind of sequence information that identified the correct reading frames. As mentioned above, the suggestion is that a simple repetitive RNY code may have served such a function in the primitive code (26, 31). It is therefore understandable that a ubiquitous RNN codon preference in present-day genes has been interpreted as a molecular fossil (87, 88). Nevertheless, this must be seen as a circular argument. Thus, the documented variability of the genetic code suggests that within the time available for the evolution of modern codes, this preference should have vanished were it not stabilized by some sort of selective pressures. Indeed, analyses of the rates of silent substitutions, the frequencies of base doublets, and synonymous codon ratios provide strong arguments that the RNN bias is stabilized in modern genes by strong selective pressures for GNN codons or, alternatively, for their amino acids (92, 100). In other words, it is difficult to argue about the form of primitive codes on the basis of modern codon preferences.

A quite plausible, though not problem-free, account of the functions of the GNN preference has been forwarded (92). In this account it is noted that there is a preference for G in the first codon position and for the avoidance of G in the second position. The ratio of the one to the other is on average close to 2, and this is observed over a very large sampling of organisms. Since different codons are responsible for this periodicity in different organisms, it is argued that the G nucleotide in the first position, and not the codons per se, is selected. These data and many others are used to support the notion that there is a universal G–non-G–N codon motif in modern mRNAs and that this statistical motif is used to monitor the reading frame of the mRNAs by the ribosome (92).

A provocative correlation is found in mRNA species that support reading frame shifts. For example, the high-frequency reading frame shift in the translation of the RF2 mRNA in *E. coli* (23, 24), as well as a number of less dramatic shifts, can be correlated with a clear sequence shift in the G–non-G–N motif from one phase to another (92). Nevertheless, it must be added that removal of the downstream motif does not seem to depress the tendency to shift phase (96) and that other specific short sequences are very clearly required to support the high-frequency shift in the RF-2 mRNA (96). Therefore, the correlation of the G–non-G–N motif with the unusual high-frequency reading frame shift is not straightforward, but at the same time it is not clear that this particular reading frame shift is a good model for normal reading frame maintenance (see above).

According to Trifonov's model, the physical connection between the G–non-G–N motif in the mRNA and the ribosome is provided by a repeat structure in the rRNA that presents a string of complementary C's with a periodicity of one in three nucleotides. Three appropriate strings have been identified in the 16S RNA, and there is good reason to believe that they are accessible to mRNA in the ribosome complex (92). Nevertheless, each of these sequences has additional C's out of phase with the once-every-third-nucleotide motif. Similarly, a device is needed to explain why looping out of nucleotides in short mRNA sequence would not disturb this phasing mechanism by presenting alternative mRNA structures to the ribosome. In summary, there are good reasons to continue to explore the functions of the ubiquitous G–non-G–N motif in reading frame maintenance. Likewise, there are equally good reasons to be skeptical about whether it is functioning as has been suggested.

Finally, the correlation that we have discussed above that associates pause sites with short strings of rare codons naturally leads to a conjecture concerning the folding process for some proteins. Thus, a rare codon string that transiently slows down the transit of the ribosome over the mRNA might facilitate proper folding of a protein domain (56, 79, 80). In particular, when the incomplete polypeptides can be arranged in alternative folds, the order of the folding could be relevant. Here, a pause site that permits the one domain to be organized before an alternative possibility is elaborated might provide a smoother assembly pathway for certain proteins. There is no strong evidence either to support or to rule out this conjecture. However, the combination of site-directed mutagenesis and protein structure studies that have become feasible now certainly offers the opportunity for a direct test of this conjecture.

CONCLUSIONS

In this review, we have emphasized the notion that the genetic code is not a frozen code. In particular, we have presented a view of evolving codon assignments as well as codon usage patterns as the adaptive response of genomes to the solution of practical problems of gene expression. For example, we have associated the evolution of the sense codon reassignments of the vertebrate mitochondria with a strategy to reduce the entire genetic apparatus of these systems. Similarly, we have associated the tendency to use only a subset of codons in the highly expressed genes of microorganisms with a growth maximization strategy. These two strategies are related in the sense that both of them are dependent on the coevolution of codon patterns and tRNA abundances. More precisely, both strategies depend on a tendency to use codons in such a way that the complexity of

tRNA populations is reduced. This similarity leads to the notion that codon reassignments are usefully viewed as extreme forms of codon preference strategies.

Our view of the evolution of codon strategies does not exclude a role for biased mutation pressure, a role that has been emphasized by others (73). However, we wish to draw attention to the selective role of a functional feedback between constraints on gene expression and the microstructure of genomes. In our view, codon usage patterns are evolving along with other characteristics of a genetic system as the result of an interplay between mutational and selective forces. The particular selective forces that we have emphasized here are directed toward a minimization of the biosynthetic apparatus of mitochondria and a maximization of microbial growth rates. There is no doubt in our minds that other such genomic strategies remain to be described.

LITERATURE CITED

1. **Andersson, D. I., H. W. van Verseveld, A. H. Stouthamer, and C. G. Kurland.** 1986. Suboptimal growth with hyper-accurate ribosomes. Arch. Microbiol. **144**:96–101.
2. **Andersson, S. G. E., R. H. Buckingham, and C. G. Kurland.** 1984. Does codon composition influence ribosome-function? EMBO J. **3**:91–94.
3. **Aota, S., and T. Ikemura.** 1986. Diversity in G+C content at the third positions of codons in vertebrate genes and its cause. Nucleic Acids Res. **14**:6345–6355.
4. **Attardi, C.** 1985. Animal mitochondrial DNA: an extreme example of genetic economy. Int. Rev. Cytol. **93**:93–145.
5. **Bennetzen, J. L., and B. D. Hall.** 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.
6. **Bernardi, G., and G. Bernardi.** 1985. Codon usage and genome composition. J. Mol. Evol. **22**:363–365.
7. **Bernardi, G., and G. Bernardi.** 1986. Compositional constraints and genome evolution. J. Mol. Evol. **24**:1–11.
8. **Bernardi, G., B. Olofsson, J. Filipski, M. Zerial, J. Salinas, G. Cuny, M. Meunier-Rotival, and F. Rodier.** 1985. The mosaic genome of warmblooded vertebrates. Science **228**:953–958.
9. **Bilgin, N., M. Ehrenberg, and C. G. Kurland.** 1988. Is translation inhibited by noncognate ternary complexes? FEBS Lett. **233**:95–99.
10. **Bilgin, N., L. A. Kirsebom, M. Ehrenberg, and C. G. Kurland.** 1988. Mutations in ribosomal proteins L7/L12 perturb EF-G and EF-Tu functions. Biochimie **70**:611–618.
11. **Bonekamp, F., H. D. Andersen, T. Christensen, and K. F. Jensen.** 1985. Codon-defined ribosomal pausing in *Escherichia coli* detected by using the *pyr E* attenuator to probe the coupling between transcription and translation. Nucleic Acids Res. **13**:4113–4123.
12. **Bonekamp, F., H. Dalböge, T. Cristensen, and K. F. Jensen.** 1989. Translation rates of individual codons are not correlated with tRNA abundances or with frequencies of utilization in *Escherichia coli*. J. Bacteriol. **171**:5812–5816.
13. **Bonekamp, F., and K. F. Jensen.** 1988. The AGG codon is translated slowly in *E. coli* even at very low expression level. Nucleic Acids Res. **16**:3013–3024.
14. **Bouadloun, F., D. Donner, and C. G. Kurland.** 1983. Codon-specific missense errors *in vivo*. EMBO J. **2**:1351–1356.
15. **Breton, R., H. Sanfacon, I. Papayannopoulos, K. Biemann, and J. Lapointe.** 1986. Glutamyl-tRNA synthetase of *Escherichia coli*. J. Biol. Chem. **261**:10610–10617.
16. **Brown, W. M., M. George, and A. C. Wilson.** 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76**:1967–1971.
17. **Bulmer, M.** 1987. Coevolution of codon usage and tRNA abundance. Nature (London) **325**:728–730.
18. **Bulmer, M.** 1988. Are codon usage patterns in unicellular organisms determined by selection-mutation balance? J. Evol. Biol. **1**:15–26.
19. **Bulmer, M.** 1988. Codon usage and intragenic position. J. Theor. Biol. **133**:67–71.
20. **Bulmer, M.** 1988. Evolutionary aspects of protein synthesis. Oxford Surv. Evol. Biol. **5**:1–40.
21. **Burns, D. M., and I. R. Beacham.** 1985. Rare codons in *E. coli* and *S. typhimurium* signal sequences. FEBS Lett. **189**:318–324.
22. **Chavancy, G., and J. P. Garel.** 1981. Does quantitative tRNA adaptation to codon content in mRNA optimize translational efficiency? Proposal for a translation system model. Biochimie **63**:187–195.
23. **Craigen, W. J., and T. Caskey.** 1986. Expression of peptide chain release factor 2 requires high-efficiency frameshift. Nature (London) **322**:273–275.
24. **Craigen, W. J., R. G. Cook, W. P. Tate, and C. T. Caskey.** 1985. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc. Natl. Acad. Sci. USA **82**:3616–3620.
25. **Crick, F. H. C.** 1968. The origin of the genetic code. J. Mol. Biol. **38**:367–379.
26. **Crick, F. H. C., S. Brenner, A. Klug, and G. Pieczenik.** 1976. A speculation on the origin of protein synthesis. Origins Life **7**:389–397.
27. **Curran, J. F., and M. Yarus.** 1988. Use of tRNA suppressors to probe regulation of *Escherichia coli* release factor 2. J. Mol. Biol. **203**:75–83.
28. **Curran, J. F., and M. Yarus.** 1989. Rates of aa-tRNA selection at 29 sense codons *in vivo*. J. Mol. Biol. **209**:65–77.
29. **Dean, A. M., D. E. Dykhuizen, and D. L. Hartl.** 1988. Fitness effects of amino acid replacements in the β-galactosidase of *Escherichia coli*. Mol. Biol. Evol. **5**:469–485.
30. **Ehrenberg, M., and C. G. Kurland.** 1984. Costs of accuracy determined by a maximal growth constraint. Q. Rev. Biophys. **17**:45–82.
31. **Eigen, M., and R. Winkler-Oswatitsch.** 1981. Transfer-RNA, an early gene? Naturwissenschaften **68**:282–292.
32. **Fast, R., T. H. Eberhard, T. Ruusala, and C. G. Kurland.** 1987. Does streptomycin cause an error catastrophe? Biochimie **69**:131–136.
33. **Fox, T. D.** 1987. Natural variations in the genetic code. Annu. Rev. Genet. **21**:67–91.
34. **Garel, J. P.** 1974. Functional adaptation of tRNA population. J. Theor. Biol. **43**:211–225.
35. **Gouy, M., and C. Gautier.** 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10**: 7055–7074.
36. **Grantham, R., C. Gautier, and M. Gouy.** 1980. Codon frequencies in 169 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res. **8**:1893–1911.
37. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.
38. **Grantham, R., C. Gautier, M. Gouy, R. Mercier, and A. Pave.** 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. **8**:49–53.
39. **Green, G. A., and D. S. Jones.** 1986. The nucleotide sequence of a cytoplasmic $tRNA^{Phe}$ from *Scenedesmus obliquus* and comparison with a $tRNA^{Tyr}$ species. Biochem. J. **236**:601–603.
40. **Grivell, L. A.** 1986. Deciphering divergent codes. Nature (London) **324**:109–110.
41. **Grosjean, H., S. de Henaue, and D. Crothers.** 1978. The physical basis for ambiguity in genetic coding interactions. Proc. Natl. Acad. Sci. USA **75**:610–614.
42. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in procaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
43. **Grosjean, H., D. Sankoff, W. MinJou, W. Fiers, and R. J. Cedergren.** 1978. Bacteriophage MS 2 RNA: a correlation between the stability of the codon:anticodon interaction and the choice of codewords. J. Mol. Evol. **12**:113–119.
44. **Hinds, P. W., and R. D. Blake.** 1985. Delineation of coding areas in DNA sequences through assignment of codon proba-

bilities. J. Biomol. Struct. Dyn. 3:543–549.
45. **Hoekma, A., R. A. Kastelein, M. Vasser, and H. A. deBoer.** 1987. Codon replacements in the *PGK1* gene of *Saccharomyces cerevisae*: experimental approach to study the role of biased codon usage in gene expression. Mol. Cell. Biol. 7:2914–2924.
46. **Holm, L.** 1986. Codon usage and gene expression. Nucleic Acids Res. 14:3075–3087.
47. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. 146:1–21.
48. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translation system. J. Mol. Biol. 151:389–409.
49. **Ikemura, T.** 1982. Correlation between the abundance of yeast tRNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. 158:573–597.
50. **Ikemura, T.** 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2:13–34.
51. **Ikemura, T., and H. Ozeki.** 1983. Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. Cold Spring Harbor Symp. Quant. Biol. 47:1087–1097.
52. **Jukes, T. H., S. Osawa, and A. Muto.** 1987. Divergence and directional mutation pressure. Nature (London) 325:668.
53. **Kimura, M.** 1983. The neutral theory of molecular evolution. Cambrige University Press, Cambridge.
54. **Kniskern, P. J., A. Hagopian, D. L. Montgomery, P. Burke, N. R. Dunn, K. J. Hofman, W. J. Miller, and R. W. Ellis.** 1986. Unusually high-level expression of a foreign gene (hepatitis B virus core antigen) in *Saccharomyces cerevisiae*. Gene 46: 135–141.
55. **Konigsberg, W., and G. N. Gordon.** 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 80:687–691.
56. **Krasheninnikov, I. A., A. A. Komar, and I. A. Adzhubei.** 1989. The role of redundancy of the genetic code in determining the mode of cotranslational protein folding. Biochemistry (Moscow) 54:187–200.
57. **Kurland, C. G.** 1987. Strategies for efficiency and accuracy in gene expression. 1. The major codon preference: a growth optimization strategy. Trends Biochem. Sci. 12:126–128.
58. **Kurland, C. G., D. I. Andersson, S. G. E. Andersson, K. Bohman, F. Bouadloun, M. Ehrenberg, P. C. Jelenc, and T. Ruusala.** 1984. Translational accuracy and bacterial growth, p. 193–203. *In* B. F. C. Clark and H. U. Petersen (ed.), Gene expression: the translational step and its control. Munksgaard, Copenhagen.
59. **Lacey, J. C., and D. W. Mullins.** 1983. Experimental studies related to the origin of the genetic code and the process of protein synthesis. A review. Origins Life 13:3–42.
60. **Lehman, N., and T. H. Jukes.** 1988. Genetic code development by stop codon takeover. J. Theor. Biol. 135:203–214.
61. **Liljenström, H., and G. von Heijne.** 1987. Translation rate modification by preferential codon usage: intragenic position effects. J. Theor. Biol. 124:43–55.
62. **Maaløe, O.** 1979. Regulation of the protein-synthesizing machinery ribosomes, tRNA, factors and so on, p. 487–542. *In* R. F. Goldberger (ed.), Biological regulation and development. Plenum Publishing Corp., New York.
63. **Matthew, B. W.** 1987. Structural basis of protein stability and DNA-protein interaction. Harvey Lect. 81:33–51.
64. **McPherson, D. T.** 1988. Codon preference reflects mistranslational constraints: a proposal. Nucleic Acids Res. 16:4111–4120.
65. **Mikelsaar, R.** 1987. A view of early cellular evolution. J. Mol. Evol. 25:168–183.
66. **Mikkola, R., and C. G. Kurland.** 1988. Media dependence of translational mutant phenotype. FEMS Microbiol. Lett. 56: 265–270.
67. **Miller, S. L.** 1986. Current status of the prebiotic synthesis of small molecules. Chem. Scr. 26B:5–11.
68. **Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA 84:166–169.
69. **Muto, A., F. Yamao, H. Hori, and S. Osawa.** 1986. Gene organization of *Mycoplasma capricolum*. Adv. Biophys. 21: 49–56.
70. **Nagyvary, J., and J. Fendler.** 1974. Origin of the genetic code: a physical-chemical model of primitive codon assignments. Origins Life 5:357–362.
71. **Ogasawa, N.** 1985. Markedly unbiased codon usage in *Bacillus subtilis*. Gene 40:145–150.
72. **Ohama, T., F. Yamao, A. Muto, and S. Osawa.** 1987. Organization of codon usage of the streptomycin operon in *Micrococcus luteus*, a bacterium with a high genomic G+C content. J. Bacteriol. 169:4770–4777.
73. **Osawa, S., and T. H. Jukes.** 1988. Evolution of the genetic code as affected by anticodon content. Trends Genet. 4:191–198.
74. **Osawa, S., and T. H. Jukes.** 1989. Codon reassignment (codon capture) in evolution. J. Mol. Evol. 28:271–278.
75. **Osawa, S., T. Ohama, T. H. Jukes, and K. Watanabe.** 1989. Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria. J. Mol. Evol. 29:202–207.
76. **Osawa, S., T. Ohama, T. H. Jukes, K. Watanabe, and S. Yokoyama.** 1989. Evolution of the mitochondrial genetic code. II. Reassignment of codon AUA from isoleucine to methionine. J. Mol. Evol. 29:373–380.
77. **Parker, J.** 1989. Errors and alternatives in reading the universal genetic code. Microbiol. Rev. 53:273–298.
78. **Pedersen, S.** 1984. *Escherichia coli* ribosomes translate *in vivo* with variable rate. EMBO J. 3:2895–2898.
79. **Pedersen, S.** 1984. In *Escherichia coli* individual genes are translated with different rates *in vivo*, p. 101–107. *In* B. F. C. Clark and H. U. Petersen (ed.), Gene expression: the translational step and its control. Munksgaard, Copenhagen.
80. **Purvis, I. J., A. J. E. Bettany, T. C. Santiago, J. R. Coggins, K. Duncan, R. Eason, and A. J. P. Brown.** 1987. The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo*: a hypothesis. J. Mol. Biol. 193:413–417.
81. **Randall, L. L., L.-G. Josefsson, and S. J. S. Hardy.** 1980. Novel intermediates in the synthesis of maltose-binding protein in *Escherichia coli*. Eur. J. Biochem. 107:375–379.
82. **Robinson, M., R. Lilley, S. Little, J. S. Emtage, G. Yarranton, P. Stephens, A. Millican, M. Easton, and G. Humphreys.** 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. Nucleic Acids Res. 12:6663–6671.
83. **Sharp, P. M., and W.-H. Li.** 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. 24:28–30.
84. **Sharp, P. M., and W.-H. Li.** 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. Nucleic Acids Res. 14:7737–7749.
85. **Sharp, P. M., and W.-H. Li.** 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4:222–230.
86. **Sharp, P. M., T. M. F. Tuohy, and K. R. Mosurski.** 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14: 5125–5143.
87. **Shepherd, J. C. W.** 1984. Fossil remnants of a primeval genetic code in all forms of life? Trends Biochem. Sci. 1:8–10.
88. **Shepherd, J. C. W.** 1986. Origins of life and molecular evolution of present-day genes. Chem. Scr. 26B:75–83.
89. **Shields, X. C., and P. M. Sharp.** 1987. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational constraints. Nucleic Acids Res. 15:8023–8040.
90. **Sörensen, M. A., C. G. Kurland, and S. Pedersen.** 1989. Codon usage determines translation rate in *Escherichia coli*. J. Mol. Biol. 207:365–377.
91. **Spanjaard, R. A., and J. van Duin.** 1988. Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. Proc. Natl. Acad. Sci. USA 85:7967–7971.

92. **Trifonov, E. N.** 1987. Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16S rRNA nucleotide sequences. J. Mol. Biol. **194**:643–652.
93. **Walker, J. E., M. Saraste, and N. J. Gay.** 1984. The *unc* operon, nucleotide sequence, regulation and structure of ATP-synthase. Biochim. Biophys. Acta. **768**:164–200.
94. **Varenne, S., J. Buc, R. Lloubes, and C. Lazdunski.** 1984. Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. **180**:549–576.
95. **Varenne, S., and C. Lazdunski.** 1986. Effect of distribution of unfavourable codons on the maximum rate of gene expression by an heterologous organism. J. Theor. Biol. **120**:99–110.
96. **Weiss, R. B., D. M. Dunn, A. E. Dahlberg, J. F. Atkins, and R. F. Gesteland.** 1988. Reading frame switch caused by base-pair formation between the 3′ end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. EMBO J. **7**:1503–1507.
97. **Wek, R. C., C. A. Hauser, and G. W. Hatfield.** 1985. The nucleotide sequence of the *ilvBN* operon of *Escherichia coli*: sequence homologies of the acetohydroxy acid synthase isozymes. Nucleic Acids Res. **13**:3995–4010.
98. **Woese, C. R.** 1965. On the evolution of the genetic code. Proc. Natl. Acad. Sci. Wash. **54**:1546–1552.
99. **Wong, J. T.** 1988. Evolution of the genetic code. Microbiol. Sci. **5**:174–181.
100. **Wong, J. T., and R. Cedergren.** 1986. Natural selection versus primitive gene structure as a determinant of codon usage. Eur. J. Biochem. **159**:175–180.
101. **Yarus, M.** 1988. A specific amino acid binding site composed of RNA. Science **240**:1751–1758.