25 Osawa, S., and Jukes, T. H., Codon reassignment (codon capture) in evolution. J. molec. Evol. *28* (1989) 271–278.
26 Osawa, S., Ohama, T., Jukes, T. H., and Watanabe, K., Evolution of the mitochondrial genetic code I. Origin of AGR serine and stop codons in metazoan mitochondria. J. molec. Evol. *29* (1989) 202–207.
27 Osawa, S., Ohama, T., Jukes, T. H., Watanabe, K., and Yokoyama, S., Evolution of the mitochondrial genetic code II. Reassignment of codon AUA from isoleucine to methionine. J. molec. Evol. *29* (1989) 373–380.
28 Osawa, S., Muto, A., Jukes, T. H., and Ohama, T., Evolutionary changes in the genetic code. Proc. Roy. Soc. Lond. B *241* (1990) 19–28.
29 Ozeki, H., Ohyama, K., Inokuchi, H., Fukuzawa, H., Kiochi, T., Sano, T., Nakahigashi, K., and Umesono, K., Genetic system of chloroplasts. Cold Spring Harbor Symp. quant. Biol. *52* (1987) 791–804.
30 Parks, T. D., Dougherty, W. G., Levings, C. S. III, and Timothy, D. H., Identification of two methoinine transfer RNA genes in the maize mitochondrial genome. Plant Physiol. *76* (1984) 1079–1082.
30a Schön, A., Kannangara, C. G., Gough, S., and Söll, D., Protein biosynthesis in organelles requires misaminoacylation of tRNA. Nature *331* (1988) 187ff.
31 Seilhamer, J. J., and Cummings, D. J., Altered genetic code in *Paramecium* mitochondria: Possible evolutionary trends. Med. gen. Genet. *187* (1982) 236–239.
32 Suyama, Y., Two-dimensional polyacrylamide gel electrophoresis analysis of *Tetrahymena* mitochondrial tRNA. Curr. Genet. *10* (1986) 411–420.
33 Wallace, D. C. W., Structure and evolution of organelle genomes. Microbiol. Rev. *46* (1982) 208–240.
34 Weber, F., Dietrich, A., Weil, J.-H., and Maréchal-Drouard, L., A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. Nucl. Acids Res. *18* (1990) 5027–5030.
35 Wolstenholme, D. R., Okimoto, R., Macfarlane, J. L., Pont, G. A., Chamberlin, H. M., Garey, J. R., and Okada, N. A., Unusual features of lower invertebrate mitochondrial genomes, in: Structure, Function and Biogenesis of Energy Transfer Systems. Eds E. Quagriello, S. Papa, F. Palmieri and C. Saccone. Elsevier, Amsterdam 1990.
36 Wakasugi, T., Ohme, M., Shinozaki, K., and Sugiura, M., Structure of tobacco chloroplast genes of tRNA Ile (CAU), tRNA Leu (CAA), tRNA Cys (GCA), tRNA Ser (UGA) and tRNA Thr (GGU): a compilation of tRNA genes from tobacco chloroplasts. Plant molec. Biol. *7* (1986) 385–392.
37 Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J., and Woese, C. R., Mitochondrial origins. Proc. natl Acad. Sci. USA *82* (1985) 4443–4447.

0014-4754/90/11-12/1117-10$1.50 + 0.20/0

Codon context

R. H. Buckingham

URA 1139 du CNRS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, Paris (France)

Summary. The analysis of coding sequences reveals nonrandomness in the context of both sense and stop codons. Part of this is related to nucleotide doublet preference, seen also in non-coding sequences and thought to arise from the dependence of mutational events on surrounding sequence. Another nonrandom context element, relating the wobble nucleotides of successive codons, is observed even when doublet preference, codon usage and bias in amino acid doublets are all allowed for. Several phenomena related to protein synthesis have been shown in vivo to be affected by the nucleotide sequence around codons. Thus, nonsense and missense suppression, elongation rate, precision of tRNA selection and polypeptide chain termination are all affected by codon context. At present, it remains unclear how these phenomena may influence the evolution of nonrandomness in the context of codons in natural sequences.
Key words. Context effects; nucleotide sequence nonrandomness; translational suppression; polypeptide chain elongation; polypeptide chain termination.

Introduction

Effects of neighbouring sequences on codon translation, usually referred to as context effects, have frequently been evoked to explain puzzling phenomena even when direct evidence of their responsibility has been weak. As will be seen below, two quite different approaches have yielded information about the significance of codon context in translation. Experimental evidence from studies both in vivo and in vitro shows that context can affect nonsense suppression, missense suppression, translational errors and frameshifting. Secondly, statistical analysis of coding sequences reveals that the context around codons is not random, and that the tendencies towards nonrandom contexts are different according to the level of expression of the genes concerned. Part of this nonrandomness has to do with mutational pressure, and shows similarities to nonrandomness in non-coding sequences. Other elements may be due to selective pressure related to translation.

Evidence from statistical analysis for constraints on codon context

Several major studies of natural coding sequences have been performed and point to significant tendencies towards nonrandom codon context. A detailed discussion of codon preferences and codon context preferences needs to be prefaced by a consideration of factors that influence the evolution of non-coding sequences, since

these inevitably affect coding sequences as well. Organisms vary widely in base composition, and in warm-blooded vertebrates the genome consists of regions (isochores) differing in base composition but internally relatively uniform[3,4,37]. Codon choice reflects firstly local base composition (Osawa et al., this issue). Superimposed on overall composition is doublet preference, a bias in the choice of neighbouring base found in non-coding as well as coding sequences[5,29,49]. In coding sequences this affects primarily 3-1 doublet frequency (i.e. the doublet composed of the third base of one codon and the first base of the following codon). This factor differs according to the organism, and is more marked in higher vertebrates, although clearly present in prokaryotes also[5]. An important mechanism leading to nonrandom doublet preferences lies in the varying probabilities of different mutational events, which at any site are themselves dependent on surrounding sequence or local DNA structure (Osawa et al., this issue). A strong bias towards the use of a subset of synonymous codons in highly expressed genes is well documented[2,27–29,35,36,40] (and Osawa et al., this issue).

An essential question to bear in mind in considering the results of statistical analyses is whether the constraints on codon contexts observed are related to translation, or originate at other levels, such as mutational mechanisms or selection acting on DNA structure, or selection based on mRNA structure or stability. Several major studies are listed in the table, with a brief description of the type of sequence bias sought and the origin of the sequences analysed.

Early attempts at statistical analysis of synonymous codon choice suggested that any proposal to account for observed preferences would have to involve constraints based on codon context[41]. This was particularly true of eukaryotic genes but was apparent in prokaryotes also. In response to the growing evidence that the translation of individual codons can be influenced by their nucleotide context, Yarus and Folley[73] analysed the nonrandomness in base occurrence over a span of 15 bases on either side of sense codons. Bias was found to be stronger in weakly expressed than in strongly expressed genes, and was qualitatively different. The adjacent upstream base was nonrandom in both gene sets. Some codons appear much more likely than others to be found in nonrandom contexts but the observation is not confined merely to a few codons. On the 3′ side, nonrandom effects extend over the three positions of the adjacent codon in the low expression set, but are restricted to the adjacent base in the highly expressed set. The observations concerning preferred downstream contexts in the low set are interpreted as a tendency towards a repeating R-YYR or Y-RRY-R pattern, and all strongly affected context positions display the qualitative bias predicted from this pattern. The authors propose that context preference in the low set reflects selection to decrease gene expression. One type of computer-generated control sequence, in addition to constraints on frame bias and codon usage, reproduces the probability of amino acid nearest-neighbours in real sequences. These controls do introduce some degree of context constraint.

The study of base preference around codons by Shpaer[59] is concerned with positions adjacent individual sense codons in *E. coli*. The preferred choice of certain synonymous codons is found to depend significantly on the nature of these bases. A search for pairwise preferences linking the third codon base with each position in the adjacent codons, taking account of amino acid pair bias in the protein sequences[26], confirmed the presence of 3-1 bias in both weakly and strongly expressed enterobacterial genes, showing that G and C are the main targets for context biases. The 3-3 bias was again much more evident in poorly expressed than well expressed genes. A 3-1 bias is also present in yeast and human genes, but is typical of doublet bias in non-reading frames and untranscribed eukaryotic sequences[29,49,60] and is therefore clearly unrelated to translation. The 3-3 bias of poorly expressed enterobacterial genes is absent from the yeast and human sequences, in spite of their strong codon preferences, a fact that leads Gouy[26] to question whether this bias is at all connected with translation.

Gutman and Hatfield[32] have explored the nonrandomness in the occurrence of codon pairs in *E. coli*. They take account of bias in the frequencies of amino acid pairs in the genes studied and show that in some cases this contributes most of the bias in codon pair frequency. A major part of their observations seems to be explicable in terms of the 3-1 and 3-3 preferences observed by others, particularly Gouy[26].

A clear relationship has been demonstrated by Hanai and Wada[33] between the bias in 3-1 doublet frequency in

Recent statistical analyses of context around sense codons

Author	Type of bias sought, organism
Yarus and Folley[73]	Frequency of bases in the 15 positions upstream and downstream of 37 sense codons (high expression set) and 61 sense codons (low expression set) treated as groups or in some cases individually; *E. coli.*
Shpaer[59]	Occurrence of upstream and downstream bases flanking sense codons in genes of high and low expression sets; *E. coli.*
Gouy[26]	Pairwise correlations between third codon base and each base of upstream and downstream and downstream codons, correcting for nonrandom amino acid pairs; *E. coli*, yeast, plasmids, man.
Gutman and Hatfield[32]	Occurrence of codon pairs, correcting for codon usage and nonrandom occurrence of amino acid pairs; *E. coli.*
Hanai and Wada[33]	Occurrence of third codon base in relation to downstream codon, preserving 3-1 doublet frequency, amino acid doublet frequency and codon usage; *E. coli.*
Bulmer[11]	Occurrence of third codon base in relation to bases in the same codon and the upstream and downstream adjacent codon; *E. coli*, yeast.

poorly expressed genes of *E. coli* and doublet preference in non-coding regions, using data for the latter from the work of Blake and Earley[5]. This is in line with the role of mutational bias suggested by Shields and Sharp[58]. They then looked for evidence of third letter context bias by comparison of coding sequences with randomised sequences respecting 3-1 doublet frequency, amino acid sequence and codon usage. Under these circumstances they found evidence of a 3-3 bias of a Y-Y or R-R type, present only in the correct reading frame, clearly related to that described by Yarus and Folley[73] and Gouy[26]. Using their control sequences conserving 3-1 bias, the remaining bias appears as a preference for a third position purine when followed by an RYR codon, and a third position pyrimidine when followed by an RRY, RYY or YRY codon. Furthermore, the preference patterns are not the same in different codon families specific to an amino acid. The authors argue that the bias is concerned with translation rather than DNA and interpret this observation in terms of interactions between tRNA molecules bound to the A- and P-sites of the ribosome. Bulmer[11] has approached the problem of the relation between context bias and translation by asking whether similar 3-1 frequencies are to be found in the complementary strand as in the coding strand. Both in *E. coli* and yeast this was shown to be the case, further suggesting that transitional mutation rates are at the origin of 3-1 nonrandom effects in poorly expressed genes.

The conclusions to be drawn from this substantial number of statistical studies may be summarised as follows. A large part of the nonrandomness present around individual codons may be described in terms of 3-1 and 3-3 doublet bias[26, 33, 73]. The doublet bias present in non-coding sequences exerts a strong influence on codon context, and largely explains the 3-1 bias observed in weakly expressed genes[11, 33, 58]. A 3-1 bias of comparable magnitude and displaying many of the same preferences is seen in highly expressed genes[26]. On the other hand, 3-3 bias is clearest in the weakly expressed genes of enterobacteria[26, 73], and persists when 3-1 bias, codon usage and bias in amino acid doublets are all allowed for[33]. This 3-3 bias is smaller in highly expressed genes, and is absent from eukaryotic genes[26]. It has been suggested that this represents selection for contexts that decrease gene expression[21, 73], but a more easily accepted interpretation is that an inherent 3-3 bias is partly overridden in highly expressed genes by the tendency to restrict codon usage to a subset of codons corresponding to the most abundant tRNA species in the cell[2, 28, 35, 36]. Accounting for this bias in poorly expressed genes poses a major unsolved problem. The less strongly a gene is expressed, the less likely it would seem that an effective selective pressure favouring, for example, translational efficiency or precision (as opposed to messenger sense) could be derived from its translation. Nevertheless, it is not clear how to explain 3-3 bias in terms other than translational.

Effects of codon context on nonsense codon suppression

The first reported effects of codon context were related to the efficiency of suppression of nonsense mutants[1, 8, 19, 20, 54, 55, 72], and it is this aspect of context effects that has received by far the most detailed experimental investigation. Feinstein and Altman[18] measured the ratio of suppression efficiencies of ochre suppressors at pairs of amber and ochre nonsense mutants at several sites in the *trp* operon, and related the varying efficiencies to the surrounding sequence. This led to the isolation of context mutants by Bossi and Roth[7] that permitted a greatly increased efficiency of suppression by *supE*, overcoming the deleterious effect of a deficiency in a post-transcriptional modification affecting the suppressor tRNA. These mutants identified the important influence of the nucleotide immediately downstream of the nonsense codon. A systematic study was undertaken of a large number of lacI amber sites in *Salmonella typhimurium*[8] and the same sites together with selected lacI opal sites in *E. coli*[45]. These data, later reexamined by Stormo et al.[63], showed that the two adjacent downstream nucleotides were important in modulating suppression efficiency. A discordant note was introduced, however, by the study of amber suppression by *supE* at different sites in T4 genes 22 and 23[17]. While confirming the influence of the adjacent 3′ nucleotide, this work pointed to important contributions from unidentified elements of context outside the immediate downstream codon.

These studies of nonsense suppression adopted the approach of comparing many different sites in a messenger for efficiency of suppression and looking for context elements that are correlated with suppression efficiency. This strategy yields precious information about non-site-specific effects, i.e. context elements that produce similar effects at all or most sites. However, by dealing with an unknown number of sequence variables this approach begs the question as to how far upstream or downstream nucleotide sequence affects suppression. Furthermore, it says nothing about context effects that may be important but which are site-specific; in other words, a particular base change (or codon change) nearby one nonsense site in a messenger may have an opposite effect (or no effect) if introduced nearby another nonsense site. Such effects would not be seen. For this reason, Buckingham et al.[10] studied the effect of introducing defined changes in the neighbourhood of a single nonsense site in a messenger. The site chosen was flanked by two serine codons; thus, the six-fold degeneracy of the genetic code for this amino acid allowed the introduction of many context changes without altering the amino acid sequence of the resulting polypeptide. While confirming the importance of the adjacent downstream nucleotide, these studies show that considerable effects may be attributed to changing the upstream adjacent nucleotide or the wobble nucleotide of the adjacent downstream codon.

Mechanism of context effects on the suppression of nonsense mutations

Nonsense suppression efficiency depends on the outcome of the competition between tRNA-mediated translation of the nonsense codon and the process of polypeptide chain termination. A major difficulty in interpreting experiments concerning the effects of context on nonsense suppression lies in the uncertainty as to whether context affects primarily the suppressor tRNA, the termination process, or both. Evidence exists for an influence on both processes. Experimental data suggesting that chain termination is sensitive to context, and furthermore that an important part of the overall context sensitivity is due to this mechanism, will be discussed in a following section. If context affected exclusively termination, then all nonsense suppressors would be expected to show the same pattern of context sensitivity. While this appears broadly to be true, an exception seems to exist in the case of ochre suppressors when suppressing amber codons by means of wobble pairing[1,7,18]. However, the most convincing evidence that context can directly affect the function of tRNA comes from experiments which show that missense codon suppression is sensitive to surrounding sequence, to be discussed in detail below. Several mechanisms have been suggested whereby tRNA function might be affected by context. The contribution to the stability of short double helices by the stacking of nonpaired 'dangling ends' has been observed both in interactions between short oligomers in solution[23] and in complex formation between tRNA molecules with complementary anticodons[31]. Stormo et al.[63] have argued that stabilization of the codon: anticodon complex by dangling ends is the major mechanism for context effects on nonsense suppression. They point out that an adjacent purine favours both phenomena, and furthermore that a pyrimidine (which itself stacks poorly) following the first dangling base should promote stacking of the latter on the codon: anticodon double helix.

One alternative explanation proposes that the protein component of the ternary complex, rather than (or in addition to) the tRNA component may be sensitive to codon context. Attention has been drawn to this possibility by the extreme case, to be discussed further below (see also the chapters of F. Caron and E. J. Murgola in this issue), presented by the mechanism in *E. coli* whereby selenocysteine is inserted at a few specific internal UGA codons.

Effects of context on the termination process

The normal process of termination at the codons UAG, UAA or UAG is still poorly understood (for reviews, see refs 13, 15). In *E. coli*, two protein release factors, RF1 and RF2, products of the genes *prfA* and *prfB*, are involved in the recognition of the stop codons, with overlapping specificity: RF1 displays specificity for both UAG and UAA, and RF2 for both UAA and UGA. A third factor, RF3, able to hydrolyse GTP, seems also to be involved, but its exact role is unclear. Termination in eukaryotes differs as far as the protein factors are concerned: the three prokaryotic factors appear to be replaced by a single protein[13,15].

The recent isolation of UGA suppressors in *E. coli* affected in the sequence of 16S RNA[46,47] has, however, revived interest in the proposal that a direct mRNA: rRNA interaction participates in the process of termination, and plays a role in the specificity of stop codon recognition (see E. J. Murgola's chapter in this issue for a full discussion).

Protein-chain termination at UAA codons in vitro, dependent on release factor RF1, has been shown to be sensitive to sequence downstream of the termination codon[24]. It appears that loosely stacked downstream sequences favour termination whereas nucleosides which encourage strong base stacking restrict release. A recent survey by Brown et al.[9] extends greatly earlier conclusions[38] concerning the preferred use of stop codons in prokaryotes and their preferred contexts. The most striking observation is a strong preference for UAA over UGA and UAG in highly expressed genes. Furthermore, a preference for uridine at the expense of adenine and cytidine is also correlated with the level of gene expression. The authors suggest that this extended codon UAAU is the most efficient stop signal in prokaryotes and is recognised by RF2.

Recent ingenious experiments by Martin et al.[43] support the notion that a significant part of the context effects on nonsense suppression may arise through the sensitivity of the termination process to context. These authors studied the relative activity of release factors RF1 and RF2 at 13 different UAA nonsense codons in the *lacI* gene in *E. coli*. Not only did this ratio vary, showing that one or both of the factors was functionally affected by context, but a correlation was demonstrated between the ratio of antisuppressor activity of the factors and the different suppressibility of the various UAA alleles. This implies that the context elements that determine suppression efficiency overlap with those that determine the relative activities of RF1 and RF2. Whether the interactions of the protein factor itself are context dependent is uncertain. An alternative explanation is provided by the isolation of nonsense suppressors that affect rRNA sequence rather than tRNA or other components of the system[46,47] (and E. J. Murgola, this volume). This suggests that direct interaction may be required between messenger and rRNA during termination, and it has been proposed that normal Watson-Crick base pairing is the basis of such an interaction. This would readily explain how messenger sequence nearby stop codons might influence the termination process; indeed, the original rRNA suppressor mutant showed evidence of context specificity.

Codon context and elongation

It is well established that the elongation cycle is not of constant length[67,68] (and references therein). Several factors contribute to the non-uniformity in elongation. Firstly, tRNA concentrations vary widely. To a first approximation, it was found possible to account for most of the pauses in polypeptide chain elongation in terms of closely grouped slowly translated codons by considering solely the concentrations of the corresponding isoaccepting species[68]. However, quite apart from the likely intervention of context effects, it has become clear that even when taking into account tRNA concentrations, codon translation times vary widely.

Three experimental approaches have been reported to compare the periods taken by the *E. coli* ribosome, in vivo, to translate different codons. The strategy of Curran and Yarus[16] provides a measure of the time taken to fill the A-site, whereas two other approaches both measure the complete elongation cycle. A pulse chase method developed by Pedersen[51] was used to show that some mRNAs from poorly expressed genes (eg. *lacI*) were translated about 30% more slowly than messengers from the highly expressed class. Messenger secondary structure is thought not to contribute to these differences[62]. Further refinement of this approach, using multiple copies of codons inserted in the *lacZ* gene, has permitted translation times to be measured for several codons. Thus, it was shown that the synonymous glutamic acid codons GAA and GAG are translated at three different rates[61], even though read by the same unique isoacceptor.

The approach adopted by Curran and Yarus[16] has been to exploit a particularly 'shifty' sequence in the gene coding for release factor RF2 in *E. coli*. Normally, this sequence brings into competition a termination event at an in frame stop codon, and a high frequency frameshift that permits continued elongation in the +1 reading frame, yielding the full length factor. Replacement of the stop codon by any of the sense codons YNN brings instead cognate tRNA selection at the sense codon into competition with the frameshift. Quantitation of the latter, if certain assumptions are accepted, provides a measure of tRNA selection time. A 25-fold range of tRNA selection time is revealed by these experiments, which means that the rate constants are spread over at least a 10-fold range. Codon preference in highly expressed genes correlates to selection rate rather than rate constant; in other words, the cell achieves high rates by means of high concentrations of the corresponding tRNAs.

Another approach to comparing codon translation rates has been to introduce different codons into the appropriate region of an attenuator leader peptide coding sequence[6]. From measurements on twelve codons, the authors find no correlation between translation time and codon preference or tRNA abundance.

It may thus be seen that data concerning individual codon translation rates, though limited, is becoming available. So far, none of these studies has been extended to measure directly the effects of changing the context around the codons under study. Few measurements from the different experimental approaches overlap sufficiently for meaningful comparisons to be made. However, little correlation is evident between the elongation cycle times seen from the attenuator studies and the tRNA selection times of Curran and Yarus[16]. Similarly, Sorensen and Pedersen[61] find CGA to be a slowly translated codon, whereas Curran and Yarus find tRNA selection for this codon amongst the fastest. This may be a reflection of the experimental difficulties in determining such parameters, or it may, as the authors suggest[61], indicate that the remainder of the elongation cycle for CGA translation is particularly long. However, it should also be borne in mind that, so far, no account has been taken of the fact that the codons under study by these diverse methods are present in different nucleotide contexts.

One set of recent experiments does specifically examine the effect on elongation rate of changing codon context. Folley and Yarus[21] sought an effect of context on the time taken to translate a ten-codon insert cloned upstream of *lacZ* in various expression vehicles. The codons of the insert were organized in two alternative fashions: firstly, so as to reproduce codon contexts typical of highly expressed genes, and secondly, by permutation of three pairs of synonymous codons, to produce contexts of low expression type. From measurements of β-galactosidase, and certain control experiments, it was deduced that the enzyme levels, via a mechanism dependent on effects of polarity, reflected the rate of insert translation. Thus, the authors conclude that the second type of insert is indeed translated more slowly than the first.

Effects of codon context on the suppression of missense mutations

The investigation of context effects on missense suppression has the attraction that the interpretation of effects, if found, is more straightforward than in the case of nonsense suppression. Thus, since only normal components of the elongation process are involved in missense suppression (to the extent, at least, that missense suppressor tRNAs can be considered 'normal'), the demonstration of significant effects holds consequences for the normal process of elongation. As in the case of nonsense suppression, comparisons of suppression efficiency at different sites in a message have been reported, and also studies in which systematic changes have been introduced around a single defined missense site in a coding region. Thus, Murgola et al.[48] compared suppression of several missense codons (GAA, GAG, AAA and AAG) at two sites, 211 and 234, in the *trpA* gene in *E. coli*. Certain combinations of codon and suppressor gave rise

to clear differences in suppression efficiency between the two sites. Some differences, notably the less efficient suppression of AAA234 than AAA211 by suppressors derived from *lysT*, could be explained by the suppressor inserting an amino acid other than glycine, required at position 234 for full activity of the *trpA* product. However, suppression by *glyV*(SuGAA/G) or *glyU*(SuGAA/G) of GAG was more efficient at site 234 than at site 211; this is clearly indicative of effects of context on the efficiency of the suppressor tRNA.

Buckingham et al.[10] have studied the effect of context on the suppression of AAA and AGA missense mutants at site 234 in the *trpA* gene, adopting an approach similar to that employed to study nonsense suppression at this site. Context effects were observed on the suppression of AAA by *glyT*(SuAAA/G), but not in the case of suppression of AGA. No obvious relation exists between the magnitude or direction of the effects on missense suppression and those on nonsense suppression. The predominant effect in nonsense suppression of the 3′ adjacent base was not observed. However, the 5′ adjacent base and the wobble base of the adjacent downstream codon influence both phenomena. It should be noted that the absence of any apparent effect of context on missense suppression may simply mean that codon reading by normal cognate tRNA and suppressor tRNA are equally affected by context changes.

It is interesting to remark that AAA is among those codons (about half of all codons) for which constraints on context have been observed in natural sequences[32, 59]. The occurrence of AGA in *E. coli* is too rare for valid statistical conclusions to be drawn. The observation that both 5′ and 3′ codons affect AAA234 missense suppression in *trpA* helps to resolve one difficulty in the interpretation of statistical studies, concerning the directionality of context effects. Thus, when constraints appear to exist downstream of a codon under study, it is difficult to determine whether these might reflect a functional interaction of the downstream codon on the upstream codon or the reverse.

Effects of codon context on mistranslation

Very little information exists concerning the influence of context on mistranslation errors, despite frequent allusions to the possibility that this may represent an important selective pressure on coding sequences[32, 44]. Carrier and Buckingham[12] showed in vitro that mistranslation as tryptophan of UGU (Cys) codons in synthetic messengers was fivefold higher when surrounded by U-rich codons than GUG codons. Shpaer has pointed out that this may be related to the apparent avoidance of UGU-UUU in natural messengers[59].

Mistranslation of UUU and UUC codons as leucine has been studied in phenylalanine-starved *E. coli* cells at two positions in the *argI* gene, coding for ornithine transcarbamylase[52]. A high frequency of misreading is seen at position 8 (normally UUC), whereas no misreading can be detected, meaning at least an order of magnitude less, at position 3 (normally UUU). Interestingly, inversion of the UUU and UUC codons by site-directed mutagenesis indicates that the difference in error level is related to the position and not to the particular codon concerned. One explanation for this is that the different context of the two positions is responsible for the levels of error occurring, though other explanations have yet to be excluded. It is possible that peptidyl-tRNA drop-off or frameshifting obscure mistranslation at position 3.

Nonsense codon readthrough as a biological strategy

Even in the absence of mutant suppressor tRNA species that read nonsense codons, polypeptide chain termination at natural stop codons is not a completely efficient process. Estimates of the degree of leakiness vary widely, consistent with the notion that natural suppression (i.e. in the absence of suppressor mutations), like that due to tRNA suppressor mutants, is sensitive to the effects of context. In general, UGA seems to be the most leaky nonsense codon (10^{-4} to $> 10^{-2}$), followed by UAG (10^{-4} to 7.10^{-3}) and then UAA (10^{-5} to 10^{-3}; see refs 8, 25, 45, 53, 56, 64). The preferred use of UAA in *E. coli* as the natural stop codon, especially in highly expressed genes[9], is perhaps related to the low leakiness of this codon.

The potential of using readthrough to produce physiologically important products has not been ignored by nature, particularly by viruses (for reviews, see refs 50, 66). Two easily distinguishable mechanisms are found, both of which may be enhanced by the selection of favourable contexts: readthrough in the sense employed above, in which the stop codon is translated as a sense codon, and frameshifting, where a fraction of the ribosomes circumvent the stop signal (for a discussion of frameshifting, see refs 70, 71).

In prokaryotic systems, it is well established that the bacteriophage Qβ produces an essential protein by partial translation of the coat protein UGA stop codon as tryptophan[69]. There is no clear evidence that *E. coli* itself uses readthrough to produce essential proteins, though it has been suggested that UGA readthrough is an important element of *trp* attenuator function[39]; it is curious, however, that natural isolates of *E. coli* frequently carry amber suppressor tRNAs[42]. Most other examples of readthrough as a biological strategy concern RNA viruses in plant or animal systems (for reviews, see refs 50, 66). An intriguing example of UGA being translated as a sense codon with high efficiency concerns the incorporation of the rare amino acid selenocysteine into formate dehydrogenase in *E. coli*[74] and glutathione peroxidase in mammals[14, 34, 65] (chapters of F. Caron and E. J. Murgola in this issue). In *E. coli*, a special mechanism exists for this alternative reading of UGA, which requires a UGA-reading tRNA (the *selC* product), and an analogue

of EF-Tu (the *selB* product) needed to recognise the unusual tRNA. Since UGA in *E. coli* is often used as a stop codon, it is clear that a particular context is required to direct selenocysteine incorporation. The essential elements of this context are unknown; nor have the components involved in recognising it so far been identified, though an attractive candidate is clearly the *selB* product itself[22]. This possibility adds credibility to the proposal that the function of normal EF-Tu may be directly affected by codon context.

Conclusions

Statistical analyses of coding sequences in *E. coli* show that striking codon context nonrandomness is present in genes from both highly expressed and poorly expressed families. This is partly derived from mutational bias and is related to the doublet bias that can be seen in non-coding sequences. Other elements of the context nonrandomness may be related to translation. It is not easy to understand how effective evolutionary pressure can be exerted through translation on genes which are poorly expressed, though it is these sequences that show most clearly the type of bias likely to depend on translation. Nevertheless, several experimentally accessible translational parameters are indeed affected by nucleotide sequence around codons. Thus, results from experiments in vivo indicate that nonsense and missense suppression, elongation rate, precision of tRNA selection and polypeptide chain termination are all affected by codon context. We cannot yet point clearly to coding sequence nonrandomness that is the result of evolutionary pressure based on these effects, except perhaps in the case of polypeptide chain termination. However, the current explosive increase in the availability of sequence data may soon allow more precise questions to be asked about context: whether, for example, identifiable bias exists in the context of codons translated by a particular isoacceptor tRNA. Furthermore, selective pressure at the level of translation may influence messenger sequence in ways difficult to detect by statistical analysis. Thus, selection against missense errors is likely to be effective only at a minority of sites in a protein, except insofar as missense errors lead to errors of processivity. Although evidence is accumulating that codon context does indeed affect the functioning of the protein synthetic machinery, the impact of these phenomena on the evolution of coding sequences remains to be fully evaluated.

1 Akaboshi, E., Inouye, M., and Tsugita, A., Effect of neighboring nucleotide sequences on suppression efficiency in amber mutants of T4 phage lysozyme. Molec. gen. Genet. *149* (1976) 1–4.

2 Andersson, S. G. E., and Kurland, C. G., Codon preferences in free-living organisms. Microbiol. Rev. *00* (1990) 00–00.

3 Aota, S.-I., and Ikemura, T., Diversity in G + C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res. *14* (1986) 6345–6355, 8702 (erratum).

4 Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meurnier-Rotival, M., and Rodier, F., The mosaic genome of warm-blooded vertebrates. Science *228* (1985) 953–958.

5 Blake, R. D., and Earley, S., Distribution and evolution of sequence characteristics in the *E. coli* genome. J. biomolec. Struct. Dynam. *4* (1986) 291–307.

6 Bonekamp, F., Dahlboge, H., Christensen, T., and Jensen, K. J., Translation rates of individual codons are not correlated with tRNA abundance or with frequencies of utilisation in *Escherichia coli*. J. Bact. *171* (1989) 5812–5816.

7 Bossi, L., and Roth, J. R., The influence of codon context on genetic code translation. Nature *286* (1980) 123–127.

8 Bossi, L., Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. J. molec. Biol. *164* (1983) 73–87.

9 Brown, C. M., Stockwell, P. A., Trotman, C. N. A., and Tate, W. P., The signal for termination of protein synthesis in prokaryotes. Nucleic Acids Res. *18* (1990) 2079–2086.

10 Buckingham, R. H., Murgola, E. J., Sörensen, P., Pagel, F. T., Hijazi, K. A., Mims, B. H., Figueroa, N., Brechemier-Baey, D., and Coppin-Raynal, E., Effects of codon context on the suppression of nonsense and missense mutations in the *trpA* gene of *Escherichia coli*, in: The Ribosome: Structure, Function and Evolution, pp. 541–545. Eds W. E. Hill, A. Dahlberg, R. A. Garret, P. B. Moore, D. Schlessinger and J. R. Warner. American Society for Microbiology, Washington, DC, 1990.

11 Bulmer, M., The effect of context on synonymous codon usage in genes with low codon usage bias. Nucleic Acids Res. *18* (1990) 2869–2873.

12 Carrier, M. J., and Buckingham, R. H., An effect of codon context in the mistranslation of UGU codon in vitro. J. molec. Biol. *175* (1984) 29–38.

13 Caskey, C. T., Peptide chain termination. Trends biochem. Sci. *5* (1980) 234–237.

14 Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W., and Harrison, P. R., The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the termination codon TGA. EMBO J. *5* (1986) 1221–1227.

15 Craigen, W. J., Lee, C. C., and Caskey, C. T., Recent advances in peptide chain termination. Molec. Microbiol. *4* (1990) 861–865.

16 Curran, J. F., and Yarus, M., Rates of aminoacyl-tRNA selection at 29 sense codons in *Escherichia coli*. J. molec. Biol. *209* (1989) 65–77.

17 Edelmann, P., Martin, R., and Gallant, J., Nonsense suppression context effects in *Escherichia coli* bacteriophage T4. Molec. gen. Genet. *207* (1987) 517–518.

18 Feinstein, S. I., and Altman, S., Context effects on nonsense codon suppression in *E. coli*. Genetics *88* (1978) 201–209.

19 Fluck, M. M., and Epstein, R. H., Molec. gen. Genet. *177* (1980) 615–627.

20 Fluck, M. M., Salser, W., and Epstein, R. H., The influence of the reading context upon the suppression of nonsense codons. Molec. gen. Genet. *151* (1977) 137–149.

21 Folley, L., and Yarus, M., Codon context from weakly expressed genes reduce expression in vivo. J. molec. biol. *209* (1989) 1–20.

22 Forchhammer, K., Leinfelder, W., and Böck, A., Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature, *342* (1989) 453–456.

23 Freier, S. M., Burger, B. J., Alkema, D., Neilson, T., and Turner, D. H., Effects of 3′ dangling end stacking on the stability of GGCC and CCGG double helices. Biochemistry *22* (1983) 6198–6206.

24 Ganoza, M. C., Buckingham, K., Hader, P., and Neilson, T., Effect of base sequence on in vitro protein-chain termination. J. biol. Chem. *259* (1984) 14 101–14 104.

25 Garen, A., and Siddiqui, O., Suppression of mutations in the alkaline phosphatase structural cistron of *Escherichia coli*. Proc. natl Acad. Sci. USA *48* (1962) 1121–1127.

26 Gouy, M., Codon contexts in enterobacterial and coliphage genes. Molec. Biol. Evol. *4* (1987) 426–444.

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

28 Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R., Codon catalogue usage is a genome strategy modulated for gene expression. Nucleic Acids Res. *9* (1981) r43–74.

29 Grantham, R., Perrin, P., and Mouchiroud, D., Patterns of codon usage of different kinds of species. Oxford Surv. evol. Biol. *3* (1986) 48–81.

30 Grosjean, H., and Fiers, W., Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene *18* (1982) 199–209.

31 Grosjean, H., Söll, D. G., and Crothers, D. M., Studies of the complex between transfer RNAs with complementary anticodons. I. Origins of enhanced affinity between complementary triplets. J. molec. Biol. *103* (1976) 499–519.
32 Gutman, G. A., and Hatfield, G. W., Nonrandom utilization of codon pairs in *Escherichia coli*. Proc. natl Acad. Sci. USA *86* (1989) 3699–3703.
33 Hanai, R., and Wada, A., Novel third-letter bias in *Escherichia coli* codons revealed by rigourous treatment of coding constraints. J. molec. Biol. *207* (1989) 655–660.
34 Hawkes, W. C., and Tappel, A. L., In vitro synthesis of glutathione peroxidase from selenite: translational incorporation of selenocysteine. Biochim. biophys. Acta *739* (1983) 225–234.
35 Ikemura, T., Correlation between abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. molec. Biol. *146* (1981) 1–21.
36 Ikemura, T., 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 *Escherichia coli* translation system. J. molec. Biol. *151* (1981) 389–409.
37 Ikemura, T., and Aota, S.-I., Global variation in G + C content along genome DNA. J. molec. Biol. *203* (1988) 1–13.
38 Kohli, J., and Grosjean, H., Usage of the three termination codons: compilation and analysis of the known eukaryotic and prokaryotic translation termination sequences. Molec. gen. Genet. *182* (1981) 430–439.
39 Kopelowitz, J., Schoulaker-Schwartz, R., Lebanon, A., and Engelberg-Kulka, H., Modulation of *Escherichia coli* tryptophan (*trp*) attenuation by the UGA readthrough process. Molec. gen. Genet. *196* (1984) 541–545.
40 Kurland, C. G., Strategies for efficiency and accuracy in gene expression. Trends biochem. Sci. *12* (1987) 126–128.
41 Lipman, D. J., and Wilbur, W. J., Contextual constraints on synonymous codon choice. J. molec. Biol. *163* (1983) 363–376.
42 Marshall, B., and Levy, S. B., Prevalence of amber suppressor-containing coliforms in the natural environment. Nature, Lond. *286* (1980) 524–525.
43 Martin, R., Weiner, M., and Gallant, J., Effects of release factor context at UAA codons in *Escherichia coli*. J. Bact. *170* (1988) 4714–4717.
44 McPherson, D. T., Codon preference reflects mistranslational constraints: A proposal. Nucleic Acids Res. *16* (1988) 4111–4120.
45 Miller, J. H., and Albertini, A. M., Effects of surrounding sequence on the suppression of nonsense codons. J. molec. Biol. *164* (1983) 59–71.
46 Murgola, E. J., Göringer, H. U., Dahlberg, A. E., and Hijazi, K. A., Ribosomal RNA and UGA-dependent peptide chain termination, in: Molecular Biology of RNA, pp. 221. Ed, T. Cech. Alan R. Liss, Inc., New York 1989.
47 Murgola, E. J., Hijazi, H. A., Göringer, H. U., and Dahlberg, A. E., Mutant 16S ribosomal RNA: a codon specific translational suppressor. Proc. natl Acad. Sci. USA *85* (1988) 4162–4165.
48 Murgola, E. J., Pagel, F. T., and Hijazi, K. A., Codon context effects in missense suppression. J. molec. Biol. *175* (1984) 19–27.
49 Nussinov, R., Doublet frequencies in evolutionally distinct groups. Nucleic Acids Res. *12* (1984) 1749–1763.
50 Parker, J., Errors and alternatives in reading the universal genetic code. Microbiol. Rev. *53* (1989) 273–298.
51 Pedersen, S., *Escherichia coli* ribosomes translate in vivo with variable rate. EMBO J. *3* (1984) 2895–2898.
52 Precup, J., Ulrich, A. K., Roopnarine, O., and Parker, J., Context specific misreading of phenylalanine codons. Molec. gen. Genet. *218* (1989) 397–401.
53 Ryden, S. M., and Isaksson, L. A., A temperature-sensitive mutant of *Escherichia coli* that shows enhanced misreading of UAG/A and increased efficiency for some tRNA suppressors. Molec. gen. Genet. *193* (1984) 38–45.
54 Salser, W., Fluck, M., and Epstein, R., The influence of reading context upon the suppression of nonsense codons, III. Cold Spring Harbor Symp. quant. Biol. *34* (1970) 513–520.
55 Salser, W., The influence of reading context upon the suppression of nonsense codons. Molec. gen. Genet. *105* (1969) 125–130.
56 Sambrook, J. F., Fan, D. P., and Brenner, S., A strong suppressor specific for UGA. Nature, Lond. *214* (1967) 452–453.
57 Sharp, P. M., and Li, W., Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res. *14* (1986) 7737–7749.
58 Shields, D. C., and Sharp, P. M., Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. Nucleic Acids Res. *15* (1987) 8023–8040.
59 Shpaer, E. G., Constraints on codon context in *Escherichia coli* genes: their possible role in modulating the efficiency of translation. J. molec. Biol. *188* (1986) 555–564.
60 Smith, T. F., Waterman, M. S., and Sadler, J. R., Statistical characterization of nucleic acid sequence functional domains. Nucleic Acids Res. *11* (1983) 2205–2220.
61 Sorensen, M. A., and Pedersen, S., Measurements of in vivo translation rates in *Escherichia coli*. 13th International tRNA Workshop, Abstr. (1989) Mo-am-13.
62 Sorensen, M. A., Kurland, C. G., and Pedersen, S., Codon usage determines translation rate in *Escherichia coli*. J. molec. Biol. *207* (1989) 365–377.
63 Stormo, G. D., Schneider, T. D., and Gold, L., Quantitative analysis of the relationship between nucleotide sequence and functional activity. Nucleic Acids Res. *14* (1986) 6661–6679.
64 Strigini, P., and Brickman, E., Analysis of specific misreading in *Escherichia coli*. J. molec. Biol. *75* (1973) 659–672.
65 Sukenaga, Y., Ishida, K., Takeda, T., and Takagi, K., cDNA sequence coding for human glutathione peroxidase. Nucleic Acids Res. *15* (1987) 71–78.
66 Valle, R. P. C., and Morch, M.-D., Stop making sense: or regulation at the level of termination in eukaryotic protein synthesis. FEBS Lett *235* (1988) 1–15.
67 Varenne, S., Knibiehler, M., Cavard, D., Morlon, J., and Lazdunski, C., Variable rate of polypeptide chain elongation for colicins A, E2 and E3. J. molec. Biol. *159* (1982) 57–70.
68 Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C., Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. molec. Biol. *180* (1984) 549–576.
69 Weiner, A. M., and Weber, K., A single UGA codon functions as a natural termination signal in the coliphage Qβ coat protein cistron. J. molec. Biol. *80* (1973) 837–855.
70 Weiss, R. B., Dunn, D. M., Atkins, J. F., and Gesteland, R. F., Slippery runs, shifty stops, backward steps, and forward hops: − 2, − 1, + 1, + 2, + 5 and + 6 ribosomal frameshifting. Cold Spring Harbor Symp. quant. Biol. *52* (1987) 687–693.
71 Weiss, R. B., Dunn, D. M., Dahlberg, A. E., Atkins, J. F., and Gesteland, R. F., 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* (1988) 1503–1507.
72 Yahata, H., Ocada, Y., and Tsugita, A., Adjacent effect on suppression efficiency, II. Study on ochre and amber mutants of T4 phage lysozyme. Molec. gen. Genet. *106* (1970) 208–212.
73 Yarus, M., and Folley, L. S., Sense codons are found in specific contexts. J. molec. Biol. *182* (1985) 529–540.
74 Zinoni, F., Birkmann, A., Stadtmann, T. C., and Böck, A., Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-lyase-linked) from *Escherichia coli*. Proc. natl Acad. Sci. USA *83* (1986) 4650–4654.

0014-4754/90/11-12/1126-08$1.50 + 0.20/0