J Mol Evol (2003) 57:694–701
DOI: 10.1007/s00239-003-2519-1

JOURNAL OF MOLECULAR EVOLUTION

Mammalian Mutation Pressure, Synonymous Codon Choice, and mRNA Degradation

Jubao Duan, Marcos A. Antezana

Department of Psychiatry, The University of Chicago, 924 East 57th Street, R-004, Chicago, IL 60637, USA

Received: 22 April 2003 / Accepted: 30 June 2003

Abstract. The usage of synonymous codons (SCs) in mammalian genes is highly correlated with local base composition and is therefore thought to be determined by mutation pressure. The usage is nonetheless structured. For instance, mammals share with *Saccharomyces* and *Drosophila* most preferences for the C-ending over the G-ending codon (or vice versa) within each fourfold-degenerate SC family and the fact that their SCs are placed along coding regions in ways that minimize the number of T|A and C|G dinucleotides ("|" being the codon boundary). TA and CG underrepresentations are observed everywhere in the mammalian genome affecting the SC usage, the amino acid composition of proteins, and the primary structure of introns and noncoding DNA. While the rarity of CG is ascribed to the high mutability of this dinucleotide, the rarity of TA in coding regions is considered adaptive because UA dinucleotides are cleaved by endoribonucleases. Here we present *in vivo* experimental evidence indicating that the number of T|A and/or C|G dinucleotides of a human gene can affect strongly the expression level and degradation of its mRNA. Our results are consistent with indirect evidence produced by other workers and with the detailed work that has been devoted to characterize UA cleavage *in vitro* and *in vivo*. We conclude that SC choice can influence strongly mRNA function and gene expression through effects not directly related to the codon–anticodon interaction. These effects should constrain heavily the nucleotide motif composition of the most abundant mRNAs in the transcriptome, in particular, their SC usage, a usage that must be reflected by cellular tRNA concentrations and thus defines for all other genes which SCs are translated fastest and most accurately. Furthermore, the need to avoid such effects genome-wide appears serious enough to have favored the evolution of biases in context-dependent mutation that reduce the occurrence of intrinsically unfavorable motifs, and/or, when possible, to have induced the molecular machinery mediating such effects to rely opportunistically on already existing motif rarities and abundances. This may explain why nucleotide motif preferences are very similar in transcribed and nontranscribed mammalian DNA even though the preferences appear to be adaptive only in transcribed DNA.

Key words: Synonymous codon — Codon usage — Mammalian — Human — Mutation — Context-dependent mutation — Noncoding region — Transcribed region — Transcriptome — tRNA pools — mRNA — Codon usage bias — UA — CG — TA — Dinucleotide — Endoribonuclease — Target sequence — Cleavage — Secondary structure — Primary structure — mRNA degradation — mRNA decay — mRNA translatability

Introduction

The fact that the major (most frequent) synonymous codons (SCs) can be read more rapidly (Grantham et al. 1981; Bennetzen and Hall 1982; Grosjean and

Correspondence to: Dr. Marcos A. Antezana, Department of Human Genetics, The University of Chicago, 920 East 58th Street, CLSC-507, Chicago, IL 60637 USA; *email:* marcos@uchicago.edu

Freirs 1982) and accurately (Precup and Parker 1987; Parker 1989) when their tRNAs are overrepresented in tRNA pools (Post et al. 1979; Post and Nomura 1979) has been central to the work on the functional advantages of biased SC usage (Kafatos et al. 1977; Fiers et al. 1978). These two translation advantages are thought to underlie the positive correlations documented in a variety of organisms between the expression level of genes and their content of major SCs. Major SCs, however, may also confer functional advantages unrelated to the codon–anticodon interaction: The folding energy *in silico* of mRNAs from a variety of organisms is lowered when one modifies the SC usage and/or the placement of SCs along the coding region (Seffens and Digby 1999). In mammals and *Drosophila*, moreover, the nucleotide motifs of the major SCs are also overrepresented across codons (e.g., N|NN, "|" being the codon boundary [Nussinov 1981a; Antezana and Kreitman 1999]). Such preferences and avoidances for across-codon nucleotide motifs can result from context-dependent mutation but at least in *E. coli* and *Drosophila* natural selection participates in creating them (Berg and Silva 1997; Antezana and Kreitman, unpublished results).

The study of preferences for dinucleotide motifs began before the DNA sequencing era (Josse et al. 1961) and so did the study of their phylogenetic distribution. In particular, the across-codon dinucleotides T|A and C|G were shown to be strongly avoided in both prokaryotes and eukaryotes, and in most multicellular organisms, respectively (see references in Setlov 1976, Russell and Subak-Sharpe 1977, and Nussinov 1981a, 1981b). The preference of some endoribonucleases for UA as a cleavage sequence (Eichler and Tatar 1980) is likely to be the cause of TA rarity (Beutler et al. 1989), a preference that has been documented *in vitro* and *in vivo* using both nontranslated RNA (Eichler and Tatar 1980; Beutler et al. 1989; Qiu et al. 1998) and mRNA (Qiu et al. 1998). No workers, however, have attempted to measure the extent to which the expression level and the degradation of a mRNA can be influenced by UA cleavage. In contrast, the rarity of C|G in mammalian genes is normally ascribed to the tendency of CG to mutate to TG (Salser 1978), which produces CAs in the opposite strand and thus a TA underrepresentation that is, however, much weaker than that observed (Duret and Galtier 2000). A DNA-level advantage for CG depletion has also been proposed, however (Lennon and Fraser 1983). Note that the aforementioned lowering of the folding energy *in silico* of mRNAs with modified SC usage and/or SC placement cannot be due to motifs longer than dinucleotides since RNA folding programs neglect (i.e., do not compute) effects of longer motifs that are not reducible to dinucleotide effects (M. Zuker, personal communication). It is, however, possible that what genes really modulate by biasing SC choice and placement is the dinucleotide content (Workman and Krogh 1999).

To date molecular evolutionists have been unable to find patterns in mammalian SC usage that are clearly attributable to natural selection (see Urrutia and Hurst 2001, Konu and Li 2002, and references therein), so that this usage is thought to be governed mainly by mutation pressure. This is sometimes misunderstood as indicating that the mammalian SC usage has no functional implications, but this is not the case since in mammalian expression systems most nonmammalian mRNAs become translatable only after their SC usage is made mammal-like (Zolotukhin et al 1996). Mammalian SC usage, even if determined by mutation pressure, appears to be functionally relevant simply because tRNA concentrations have evolved to match it (Zhou et al. 1999). This, however, explains neither why specific SCs have become major SCs nor why many SCs are used in the same order of preference by distantly related organisms (Antezana and Kreitman 1999). The stability of the mRNA could be the reason but no studies have addressed this possibility.

In this paper we report that changes in the SC usage that maximize the T|A, C|G, or the T|A + C|G content of a human gene drastically affect the steady-state level and the degradation rate *in vivo* of its mRNA. The gene we chose encodes the 443 amino acids of the long isoform of the dopamine receptor D2 of humans and, like most human genes, has extremely low T|A and C|G contents. A DRD2 sequence with shuffled SC placement but unchanged SC usage and encoded amino acid sequence has probabilities of about 0.0004 and < 0.0001 of having an equal or lower count of T|A and C|G than the wild-type DRD2 sequence, respectively (estimated over 100,000 randomized sequences), i.e., the two dinucleotides are extremely underrepresented relative to what one would expect given the gene's SC usage and the amino acid sequence it encodes. The actual count and the average count of the two dinucleotides in shuffled sequences are 6 (15.5) and 36 (57.6), respectively.

Laboratory Methods

Cloning, Transfection, and Cell Culture

A 2.5-kb cDNA fragment with the long isoform of the human DRD2 coding region and 5′- and 3′-UTRs (untranslated regions) was subcloned into a pCMV-Script plasmid (Stratagene). The three modified DRD2 sequences were synthesized chemically by GeneArt (Regensburg, Germany); two natural restriction sites in the 5′- and 3′-UTRs that were very close to the coding sequence were used to subclone the three modified DRD2 sequences into the same plasmid. The four DRD2 constructs were excised from the pCMV plasmids and inserted into the *Not*I/*Xho*I site of the Flp-In expression vector pCDNA5/FRT (Invitrogen). These

constructs were then transfected into CHO host cells and integrated in the cells' only Flp-In insertion site. The Flp recombinase necessary for the integration of the vectors containing the four DRD2 sequences was supplied by cotransfecting the pOG44 plasmid (Invitrogen), which expresses this recombinase constitutively. We used the LipoTAXI mammalian transfection kit (Stratagene). Cells carrying the integration were selected in a medium containing 600 μg/ml Hygromycin B. The selected colonies (>200 per construct) were pooled by trypsinization for further study. Cells were cultured in Ham's F-12 medium (from American Type Culture Collection, Manassas, VA) supplemented with 10% fetal bovine serum at 37°C in a humidified atmosphere of 5% CO_2.

mRNA Work

An RNase protection assay was used for the detection and quantitation of DRD2 mRNA in transfected cells. We used the Direct Protect Lysate RPA kit (Ambion), and the expression level of the endogenous 28S rRNA was used as the normalizing reference. Forty microliters of cell lysate in lysis/denaturing solution from 5×10^5 cells was directly used in each RPA reaction, and 5 μl of ^{32}P-labeled DRD2 RNA probe and 5 μl of ^{32}P-labeled 28S rRNA probe (see below) were added to each reaction. The experimental procedures were those recommended by Ambion. The intensity of the bands was quantitated with a Typhoon 8600 phosphoimager (Amersham Biosciences). The data were analyzed with the PRISM 3.0 program (GraphPad). The RNA probes used in the RNase protection assays were synthesized using the MAXIscript T7 kit (Ambion) and labeled with [α-^{32}P]UTP (Amersham Biosciences). A 184-bp-long PCR-generated fragment homologous to a region of the DRD2 3′ UTR, to which a T7 promoter was added with the help of the No-Cloning Promoter Addition Kit (Ambion), was used to generate the DRD2 antisense RNA probe. The plasmid pTRI-RNA 28S (Ambion) was used to generate the 173-base long antisense 28S rRNA probe. After the RNase protection assay the expected sizes were about 174 and 115 bp for the DRD2 and the 28S rRNA fragments, respectively, i.e., they were clearly distinguishable from each other. ActD (10 mg/ml) was applied to cell cultures grown overnight, cells were harvested at selected intervals of up to 24 h, and their DRD2 mRNA content was quantitated as described above.

Experiment and Results

To produce the C|G-rich sequence every "non-C"|G motif that could become C|G without altering the amino acid sequence was so changed. This gave 52 new C|G's, for a total of 88. The T|A-rich sequence was produced through 114 synonymous changes from "non-T" to T, which raised the T|A count to 120. In addition to the aforementioned changes, the T|A + C|G-rich sequence also carried 9 T|A's produced through changes within sixfold-degenerate SC families (no such change could contribute to the C|G-rich sequence). The changes we introduced in the DRD2 coding region were quite evenly distributed, e.g., the longest tract between two successive, altered SCs in the U|A + C|G-rich sequence, was 36 bases long (Fig. 1), so that the primary and secondary structure was altered in all but the shortest stretches of DNA in the DRD2 coding region.

The steady-state level and the decay of the four different DRD2 mRNAs were measured in Chinese Hamster Ovary (CHO) cells. These cells do not express their endogenous DRD2 gene and are commonly used to study the expression and regulation of foreign DRD2 genes. Through transfection and site-specific recombination into an identical and unique chromosomal location, we integrated in the genome of CHO cells one and only one expression construct containing the intronless sequence of the wild-type human DRD2 gene or of each of the three aforementioned modified versions of the gene (see Laboratory Methods). All constructs had the same promoter sequences. Therefore differences in transcription rate due to chromosomal position effects cannot have biased the comparisons. We expected the U|A-rich and possibly the U|A + C|G-rich mRNAs to have a lower stability but had no expectations for the C|G-rich mRNA.

We first measured the expression level of DRD2 mRNA in the transfected CHO cell lineages after they reached transcription/degradation steady state. (see Laboratory Methods). We observed conspicuous differences (Fig. 2A): While the C|G-rich mRNA had an expression level 27% higher than the wild type, the levels of the U|A-rich and U|A + C|G-rich mRNAs were only 17 and 25% of that of the wild type, respectively ($p = 0.0284$, 0.0007, and 0.0007; t test). The slightly higher level of the U|A + C|G-rich mRNA relative to that of the U|A-rich mRNA was also significant ($p = 0.0188$). Obviously at least some of the 114 SC changes present in the U|A-rich mRNA caused a drastic decrease in the steady-state expression level, an effect that overwhelmed that of the changes which increased the expression level of the C|G-rich mRNA. Because every tested DRD2 sequence was integrated at the same chromosomal location and was under the control of identical promoters, the expression level differences indicated that mRNA degradation differences had shifted the transcription/degradation equilibrium.

We therefore determined the degradation rates of the four mRNAs. To this end, we measured the level of the four DRD2 mRNAs at different time points after blocking transcription (see Laboratory Methods). We found that the U|A-rich and the U|A + C|G-rich mRNAs decayed much more rapidly than the wild-type and the C|G-rich mRNAs (Fig. 2B). The half-life of U|A-rich mRNA was significantly shorter than that of the wild type, changing from 8–9 h to about 3 h. The half-life of the C|G-rich mRNA was significantly longer than that of the wild type ($p = 0.0007$ and $p < 0.0001$, respectively; two-way ANOVA). The half-life of the U|A-rich mRNA was shorter than that of the U|A + C|G-rich mRNA but not significantly so ($p = 0.2637$). Therefore, the observed differences in mRNA decay are fully consistent with the differences in steady-state expression

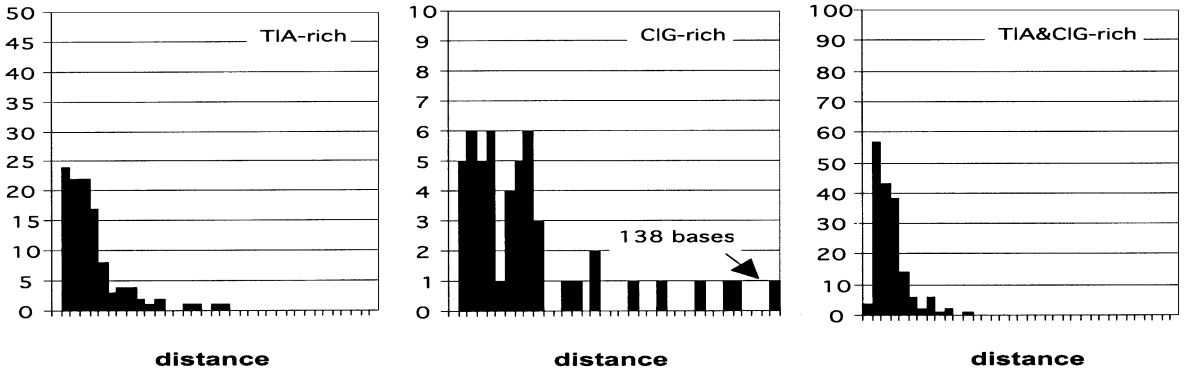

Fig. 1. The distribution of the distance (in bases) between the closest altered codons in the three synthetic, 1332-base-long DRD2 coding regions.

level documented above, indicating that the differences in steady-state mRNA levels shown in Fig. 2 are at least partially due to the four mRNAs having different degradation rates.

Discussion

The faster degradation of our UA-rich and UA + CG-rich mRNAs is consistent with the possibility that endoribonucleolytic cleavage of the added UA dinucleotides accelerated degradation. Previous workers have documented and characterized in great detail the cleavage of UAs both *in vitro* and *in vivo* using nontranslated RNA and an engineered mRNA (Eichler and Tatar 1980; Beutler et al. 1989; Qiu et al. 1998). The latter workers, in particular, showed that an RNA sequence delimited by three UAs on each side began being cleaved within minutes out of an mRNA molecule precisely where the UAs were located. This cleavage time is fully compatible with the times of degradation of our mRNAs. Further evidence consistent with the possibility that UA cleavage accelerated the degradation of our two UA-rich mRNAs can be found in three reports of mRNA stability changes that were ascribed to exchanges of minor and major SCs. In all three cases and apparently unbeknownst to the authors, the introduced base changes increased or decreased the UA and/or CA count (CA is the other target dinucleotide [Eichler and Tatar 1980]), in full agreement with the possibility that additional UAs and CAs accelerate mRNA degradation (Hoekema et al. 1987; Perlak et al. 1991; Deana et al. 1996). Perlak et al., in particular, compared two modified mRNAs to an unmodified original. The two modified mRNAs had UA + CA(UA,CA) counts of −28 (−25, −3) and −39 (−91, + 52) relative to the unmodified mRNA, and their mRNA level increased stepwise and remarkably linearly to 2.5 and 5 times that of the unmodified mRNA, respectively. All of the above supports our view that the faster degradation of our two UA-rich mRNAs was due to the fact that their additional UAs were cleaved by an endoribonuclease(s). Note, however, that the normal degradation of mammalian mRNA should not involve UA cleavage at any appreciable rate because human mRNAs have very few UAs that, moreover, should be shielded from endonucleases by secondary structure (see below). Since SC usage and SC placement along the coding region are the main determinants of the U|A count of a coding region, our results suggest that SC choice could be constrained heavily by the need to minimize the number of motifs like UA that affect mRNA degradation in undesirable ways, i.e., by the need to guarantee an acceptable steady-state expression level.

We believe that the C|G-rich mRNA was degraded more slowly because its additional CGs reinforced the secondary structure of the mRNA. Secondary structure can hinder the action of endoribonucleases (Eichler and Eales 1983) and in vivo evidence suggests that this is also so for the endonuclease(s) responsible for UA cleavage. Qiu et al. (1998) showed that an RNA sequence delimited by three UAs on each side was excised more slowly out of an RNA molecule when the length of the termini was increased. Thus it is likely that the UA + CG-rich mRNA decayed more slowly than the UA-rich mRNA because some of its numerous UAs were made less susceptible to cleavage by the increase in secondary structure created by the additional hydrogen bonds of the Cs and Gs that were introduced to create CGs and by the self-complementarity of the CG motifs themselves. It is also possible that the CG-rich mRNA decayed more slowly than the wild type for the same reason, but in this case the increase in secondary structure could only improve the shielding of the mere six wild-type UAs. Similarly, the U's that increased the UA count of the U|A-rich mRNAs may have contributed to the faster decay of these mRNAs by weakening the secondary structure. The reading frame-independent secondary-structural consequences of the base composition and of the nucleotide-motif composition should also constrain SC usage and placement over the whole coding region. The

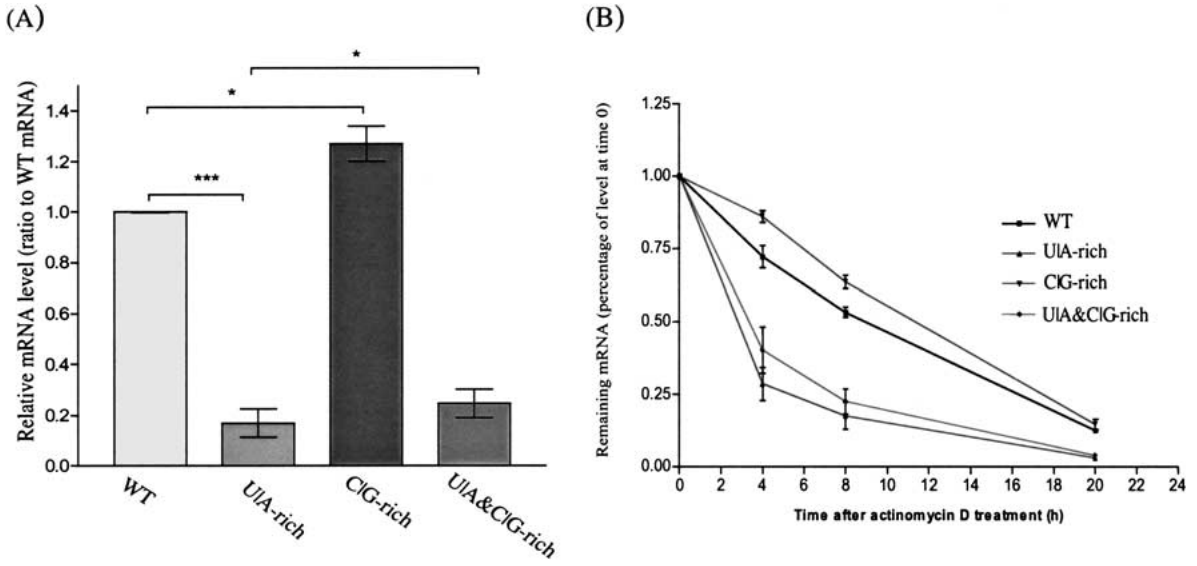

Fig. 2. The steady-state level (**A**) and decay (**B**) of three modified DRD2 mRNAs that were either U|A-rich, C|G-rich, or U|A + C|G-rich (see text), contrasted with those of the wild-type mRNA (WT). At least three independent measurements were performed for each mRNA and time point. $*p < 0.05$; $**p < 0.001$ (t test).

G + C content is an example of a base-composition feature that should affect mRNA secondary structure and may be constrained for this reason. One should remember, furthermore, that gains in mRNA stability that are due to stronger secondary structure are not necessarily beneficial because secondary structure can affect translatability (Candelas et al. 1983; Zama 1997). Therefore CGs may be deleterious even as they increase mRNA stability. Finally, effects due to those nucleotide motifs other than U|A's and C|G's, whose counts were altered when the U|A and/or the C|G count were increased, may have also contributed to our results.

The conclusion that SC usage and placement are constrained by their effects on mRNA stability and turnover would be invalid if our results were due to just a few of the changes that we introduced to the coding region of our gene and/or if most SC changes that alter nucleotides and nucleotide motifs that we did not modify were to have only minor functional consequences. The first possibility is realistic: Some of the changes to C may have disrupted a degradation element in the coding region (and/or may have improved an antidegradation one), and some of those to U may have done the contrary but more strongly so (since their effect dominated that of the introduced Cs in the U|A + C|G-rich mRNA). We consider this scenario to be unlikely, however. First, for such an alternative explanation to preempt our explanations, one would have to postulate that the demonstrated *in vivo* cleavage of UAs in nontranslated RNAs and in Qiu et al.'s (1998) engineered mRNA, and the modulation of UA cleavage by secondary structure discussed above, do not to apply to "native" mRNA like ours. Furthermore, for such an alternative explanation to apply the following appears necessary. (i) Since the UA-rich mRNA was degraded faster, at least some of the changes to UA would have to damage stability element(s) and/or "worsen" instability element(s) located in the coding region, more than the other UA changes would improve and neutralize, respectively, the same elements and/or others; (ii) Since the CG-rich mRNA was degraded more slowly, at least some of the changes to CG would have to improve stability element(s) and/or "neutralize" instability element(s) located in the coding region and would have to damage them more than the other CG-changes would disrupt and "worsen," respectively, the same elements and/or others; (iii) Since the UA + CG-rich mRNA did not decay faster than the UA-rich one, the elements "worsened" by UA changes would have to be able to tolerate the numerous CG changes that were introduced without accelerating decay additionally, or they would have to be located in the rare zones of the coding region in which no changes to CG were made (Fig. 1); (iv) Similarly, since the UA + CG-rich mRNA did not decay faster than the UA-rich mRNA, the element(s) "improved" by CG changes must be able to tolerate the changes to UA without becoming instability elements or must be located in the even rarer segments of the coding region in which no changes to UA were made (Fig. 1); (v) Finally, since the UA + CG-rich mRNA decayed only slightly more slowly than the UA-rich mRNA, the effects of those CG changes that "bettered" elements must be overwhelmed by the effects triggered by those UA changes that "worsened" elements. To the best of our knowledge, no elements have been described that intact and/or disrupted could account for our ob-

servations. Moreover, two degradation elements whose minor SCs have received much attention contain UA and/or CA dinucleotides that have apparently not been studied but could very well be important. In one case three incremental deletions of an element increased mRNA half-lives from 5 to 7.5, 35, and 45 min (Parker and Jacobson 1990), but the deletions also removed two CAs, four additional CAs and two UAs, and one additional UA, respectively. And in a case in which an element was studied by changing minor codons into major ones (Lemm and Ross 2002), the one such change that had a strong effect eliminated one of three contiguous CAs (of the four CAs in the element). All of the above, together with the body of work proving that minor SCs are not sufficient to accelerate mRNA degradation (see, e.g., the chapter in Caponigro and Parker 1996), makes it unlikely that the real cause of the faster degradation of our two UA-rich mRNAs was disruptions of putative stability elements and/or the increase in minor SCs that was necessary to raise the U|A count. Moreover, the above indicates that some instability elements could be clusters of target dinucleotides and minor codons: The minor SCs would stall translation and thus alter mRNA folding for longer periods of time, increasing the probability that target dinucleotides located closeby be exposed and cleaved (Parker and Jacobson 1990). In the absence of such dinucleotides the minor SCs would not suffice to accelerate mRNA decay, and in the absence of rare codons the target dinucleotides would be cleaved more slowly, if at all.

The second caveat above was that SC changes which do not affect U|A and C|G counts (or the |*n*UA|, |UA*n*|, |*n*CG|, and |CG*n*| counts) may not affect mRNA function greatly so that some features of SC choice could be less constrained and thus could be freer to change. The across-codon dinucleotides whose counts we maximized are indeed the most strongly underrepresented among the 16 dinucleotides, i.e., they are the dinucleotides most likely to have some large functional impact. However, one should note that (1) the two main effects documented above resulted from SC changes in only about 20 and 10% of all third positions, respectively, and (2) the largest effect above resulted from the largest primary-structural modification. Therefore, it would be not surprising if changing, say, 70% of the third positions in ways that modify the counts of motifs that are less strongly avoided or preferred would result in mRNA stability consequences comparable to those documented here. This would let most third-position bases have functional consequences, at least when acting together with other such bases, creating a whole-coding region constraint on SC.

The above mechanistic interpretations of our results are plausible because they are fully consistent both with the aforementioned work that inadvertently measured the mRNA degradation consequences of comparable primary-structural manipulations and with the aforementioned experimental work that has demonstrated and characterized thoroughly the cleavage of UAs and CAs in translated and non-translated RNA. Nonetheless, exacting laboratory work will be necessary to characterize in detail the mechanisms involved and to develop a quantitative understanding of how the endoribonucleolytic cleavage of target dinucleotides can be modulated by mRNA secondary structure, translation speed, and pauses, as well as by the proximity of minor SCs, and of how all of this participates in determining mRNA degradation rates. The study of the possible functional consequences of other preferred or avoided nucleotide motifs may turn out being very challenging as well, since some of these motifs may have subtler consequences. CA dinucleotides, in particular, are a mystery since in mammalian genes they are quite numerous, despite their having been shown to be cleaved by endoribonuclease(s), albeit so far only in special situations such as when flanking a UA (Qiu et al. 1998) and in runs (Eichler and Tatar 1980).

For the genes whose mRNAs dominate the transcriptome, specifically those whose mRNAs are needed in very high quantities, mRNA stability is undoubtedly at a premium. The magnitude of the effects documented above indicates that the consequences of SC choice for the stability of these mRNAs are likely to be at least as important as its consequences for translation speed and accuracy that are mediated by the codon–anticodon interaction (Chavancy and Garel 1981). Therefore the central constraint behind the choice of major and minor SCs, and thus behind tRNA concentration differences, might be the stability requirements of the mRNAs that are highly expressed and the optimization of other aspects of the function of these mRNAs that depend on mRNA primary structure in a reading frame-independent manner, i.e., in ways not mediated by the codon–anticodon interaction. In addition to mRNA stability, another such aspect may be translatability since translation speed can be influenced by mRNA secondary structure (Candelas et al. 1983; Zama 1997) and, possibly, by the high-energy G-quartet structures (Williamson 1994) that tend to arise in runs of Gs in any reading frame.

Under the above scenario the biased concentrations of tRNAs would be a secondary adaptation to the preferences of highly expressed genes for those SCs whose nucleotide motifs contribute best to frame-independent mRNA properties. By matching these preferences, tRNA concentrations would enhance the translation of the mRNA of highly expressed genes, i.e., enhance the vast majority of the codon–anticodon interactions that take place in the

cell. Overrepresented tRNAs would then come under pressure to improve in additional ways the codon–anticodon interaction(s) in which they participate, for example, through posttranscriptional modification of and/or primary-structural changes in their anticodon stems (Lustig et al. 1993). Genomewide genes that are less highly expressed but in need of fast and/or accurate translation would have to adopt the same SC usage. Under this scenario, the SC usages of different species could not become very different because the molecular machinery determining the consequences of the motifs of the various SCs for frame-independent mRNA properties is likely to be hard to modify and thus less prone to diverge evolutionarily, and also because an optimized tRNA participation in a codon–anticodon interaction may be an adaptation that should take some time to evolve, unlike a change in the concentration of a tRNA.

And what about the evolutionary forces structuring mammalian SC usage? On the one hand, molecular evolutionists have had to conclude that mutation pressure is the main determinant of this usage after failing to find usage patterns that depart clearly from what mutation pressure alone should produce (Urrutia and Hurst 2001; Konu and Li 2002). A mammalian-like SC usage is necessary for efficient translation in mammalian systems (Zolotukhin et al. 1996) but this appears to be a simple consequence of tRNA concentrations having adapted to the mutation-determined SC usage, thereby making the tRNA pools less suitable for translating mRNAs with a different SC usage (Zhou et al. 1999). On the other hand, mammalian SC usage, the placement of SCs along mammalian genes, and the amino acid composition of the proteins these genes encode are biased in ways that minimize the number of TA dinucleotides anywhere in the coding region (Nussinov 1981b). TAs are therefore rare as |TA*n*|'s, |*n*TA|'s, and T|A's, exactly as one would expect if TAs are avoided because as UAs they accelerate mRNA degradation in any reading frame. However, it seems implausible mechanistically that the reaction of the mRNA degradation machinery to the various nucleotide motifs and, more generally, the consequences of the different motifs for frame-independent mRNA properties may have been modified to let the motif composition produced by mutation pressure become functionally favorable. However, when the effects are mediated by molecular machinery that in principle could react to any specific motif (e.g., by choosing the motif as target sequence), the machinery could exploit opportunistically motifs rarities or abundances caused by spontaneous mutational biases. In the case of TA, the latter may have happened early in the history of life since TA underrepresentations are found almost universally (Setlov 1976, Russell and Subak-Sharpe 1977; Nussinov 1981b). An alternative explanation would be that the mutation pressure has been tuned by natural selection to remove from the DNA those motifs that are detrimental to frame-independent mRNA properties. Adaptive mutation is anathema to evolutionists, but in this case it could have evolved easily as a result of simple and strong selection (1) to reduce the mutation rate in transcribed DNA toward motifs like TA that can be very deleterious to the mRNA in any reading frame even in low numbers and (2) to reduce the rate at which favorable motifs mutate to less favorable motifs. This would in turn result in "higher" (less curbed) mutation rates toward less deleterious motifs, especially if trade-offs limit the capability of replication and repair enzymes to lower the mutation rate in every primary-structural context. Direct evidence in this direction comes from studies of somatic hypermutation showing that TAs (and CGs) mutate more often than other motifs (Smith et al. 1996) and that hypermutation and meiotic mutation have similar context preferences (Oprea et al. 2001). If biased context-dependent mutation were not the main cause of TA rarity and of the other motif over- and underrepresentations typical of the mammalian genome, including and especially of those observable in its 3000 Mb of noncoding DNA (Nussinov 1981b; Karlin and Burge 1995), one would have to postulate that natural selection protects these preferences from being erased by mutation. This, however, is implausible from both a molecular biology and a population genetics perspective. It is indeed unlikely that selection may be lowering TA counts to a similar extent over the whole genome because a TA count serendipitously similarly low happens to be a panacea adaptation that improves the stability of the mRNA, the primary structure of introns, unknown functional aspects of noncoding DNA, etc. And it is also unlikely that any mammalian population, let alone the human one, could tolerate the immense genetic load that selection would create if it had to protect against mutation the nucleotide motif composition of 3000 Mb of noncoding DNA.

Acknowledgments. We thank Elliott Gershon, Marry Kreitman, Carlos C Machado, Ira Wool, Matthew Meselson, and an anonymous reviewer for comments, advice, literature indications, and/or encouragement; and Janice Spofford, Martha Hamblin, Esther Betran, Lisa Noble, Donovan Conrad, Antonio Grinbergo, Charles Langley, Wolfgang Stephan, Peter Stadler, Mactias Gerberding, and Shin-Han Hui for comments and/or editorial advice. The synthetic sequences were financed by a grant from The University of Chicago's Brain Research Foundation to M.A. Antezana and Pablo Gejman. J. Duan was supported by a NARSAD Young Investigator Award. We warmly thank P. Gejman for his enthusiastic belief in the project, for the use of his laboratory and supplies, and for salary support.

References

Antezana MA, Kreitman M (1999) The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences. J Mol Evol 49(1):36–43

Antezana MA, Kreitman M. Unpublished data

Bennetzen JL, Hall BD (1982) Codon selection in yeast. J Biol Chem 257:3026–3031

Berg OG, Silva PJ (1997) Codon bias in Escherichia coli: The influence of codon context on mutation and selection. Nucleic Acids Res 25(7):1397–1404

Beutler E, Gelbart T, Han JH, Koziol JA, Beutler B (1989) Evolution of the genome and the genetic code: Selection at the dinucleotide level by methylation and polyribonucleotide cleavage. Proc Natl Acad Sci USA 86(1):192–196

Candelas G, Candelas T, Ortiz A, Rodriguez O (1983) Translational pauses during a spider fibroin synthesis. Biochem Biophys Res Commun 116(3):1033–1038

Caponigro G, Parker R (1996) Mechanisms and control of mRNA turnover in *Saccharomyces cerevisiae*. Microbiol Rev 60(1):233–249

Chavancy G, Garel JP (1981) Does quantitative tRNA adaptation to codon content in mRNA optimize the ribosomal translation efficiency? Proposal for a translation system model. Biochimie 63(3):187–195

Deana A, Ehrlich R, Reiss C (1996) Synonymous codon selection controls in vivo turnover and amount of mRNA in Escherichia coli bla and ompA genes. J Bacteriol 178(9):2718–2720

Duret L, Galtier N (2000) The covariation between TpA deficiency, CpG deficiency, and G + C content of human isochores is due to a mathematical artifact. Mol Biol Evol 17(11):1620–1625

Eichler DC, Eales SJ (1983) The effect of RNA secondary structure on the action of a nucleolar endoribonuclease. J Biol Chem 258(16):10049–10053

Eichler DC, Tatar TF (1980) Properties of a purified nucleolar ribonuclease from Ehrlich ascites carcinoma cells. Biochemistry 19(13):3016–3022

Fiers W, Contreras R, Haegeman G, Rogiers R, Van de Voorde A, Van Heuverswyn H, Van Herreweghe J, Volckaert G, Ysebaert M (1978) Complete nucleotide-sequence of SV40 DNA. Nature 273:113–120

Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res 9(1):R43–R74

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

Hoekema A, Kastelein RA, Vasser M, de Boer HA (1987) Codon replacement in the PGK1 gene of Saccharomyces cerevisiae: Experimental approach to study the role of biased codon usage in gene expression. Mol Cell Biol 7(8):2914–2924

Josse J, Kaiser AD, Kornberg A (1961) Enzymatic synthesis of deoxyribonucleic acid. 8. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. J Biol Chem 236(3):864–875

Kafatos FC, Efstratiadis A, Forget BG, Weissman SM (1977) Molecular evolution of human and rabbit beta-globin mRNAs. Molecular evolution of human and rabbit beta-globin messenger-RNAs. Proc Natl Acad Sci USA 74(12):5618–5622

Karlin S, Burge C (1995) Dinucleotide relative abundance extremes: A genomic signature. Trends Genet 11(7):283–290

Konu O, Li MD (2002) Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents. J Mol Evol 54(1):35–41

Lemm I, Ross J (2002) Regulation of c-myc mRNA decay by translational pausing in a coding region instability determinant. Mol Cell Biol 22(12):3959–3969

Lennon GG, Fraser NW (1983) CpG frequency in large DNA segments. J Mol Evol 19(3–4):286–288

Lustig F, Boren T, Claesson C, Simonsson C, Barcisszewska M, Lagerkvist U (1993) The nucleotide in position 33 of the tRNA anticodon loop determines ability of anticodon UCC to discriminate among glycine codons. Proc Natl Acad Sci USA 90:3343–3347

Nussinov R (1981a) Eukaryotic dinucleotide preference rules and their implications for degenerate codon usage. J Mol Biol 149(1):125–131

Nussinov R (1981b) Nearest neighbor nucleotide patterns. Structural and biological implications. J Biol Chem 256(16):8458–8462

Oprea M, Cowell LG, Kepler TB (2001) The targeting of somatic hypermutation closely resembles that of meiotic mutation. J Immunol 166(2):892–899

Parker J (1989) Errors and alternatives in reading the universal genetic code. Microbiol Rev 53(3):273–298

Parker R, Jacobson A (1990) Translation and a 42-nucleotide segment within the coding region of the mRNA encoded by the MAT alpha 1 gene are involved in promoting rapid mRNA decay in yeast. Proc Natl Acad Sci USA 87(7):2780–2784

Perlak FJ, Fuchs RL, Dean DA, McPherson SL, Fischhoff DA (1991) Modification of the coding sequence enhances plant expression of insect control protein genes. Proc Natl Acad Sci USA 88(8):3324–3328

Post LE, Nomura M (1979) Nucleotide sequence of the intercistronic region preceding the gene for RNA polymerase subunit alpha in Escherichia coli. Biol Chem 254(21):10604–10606

Post LE, Strycharz GD, Nomura M, Lewis H, Dennis PP (1979) Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in Escherichia coli. Proc Natl Acad Sci USA 76(4):1697–1701

Precup J, Parker J (1987) Missense misreading of asparagine codons as a function of codon identity and context. Biol Chem 262(23):11351–11355

Qiu L, Moreira A, Kaplan G, Levitz R, Wang JY, Xu C, Drlica K (1998) Degradation of hammerhead ribozymes by human ribonucleases. Mol Gen Genet 258(4):352–362

Russell GJ, Subak-Sharpe JH (1967) Similarity of the general designs of protochordates and invertebrates. Nature 266:533–536

Salser W (1978) Globin mRNA sequences: Analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp Quant Biol 42(2):985–1002

Seffens W, Digby D (1999) mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. Nucleic Acids Res 27(7):1578–1584

Setlow P (1976) In: Fasman GD (ed) Handbook of biochemistry and molecular biology. CRC Press, Cleveland, OH, pp 312–318

Smith DS, Creadon G, Jena PK, Portanova JP, kotzin BL, Wysocki LJ (1996) Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. J Immunol 156(7):2642–2652

Urrutia AO, Hurst LD (2001) Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics 59(3):1191–1199

Williamson JR (1994) G-quartet structures in telomeric DNA. Annu Rev Biophys Biomol Struct 23:703–730

Workman C, Krogh A (1999) No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. Nucleic Acids Res. 27:4816–4822

Zama M (1997) Translational pauses during the synthesis of proteins and mRNA structure. Nucleic Acids Symp Ser 37:179–180

Zhou J, Liu WJ, Peng SW, Sun XY, Frazer I (1999) Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. J Virol 73:4972–4982

Zolotukhin S, Potter M, Hauswirth WW, Guy J, Muzyczka N (1996) A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. J Virol 70:4646–4654