

# Amino Acid Exchangeability and the Adaptive Code Hypothesis

Arlin Stoltzfus · Lev Y. Yampolsky

Received: 30 January 2007 / Accepted: 23 August 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** Since the genetic code first was determined, many have claimed that it is organized adaptively, so as to assign similar codons to similar amino acids. This claim has proved difficult to establish due to the absence of relevant comparative data on alternative primordial codes and of objective measures of amino acid exchangeability. Here we use a recently developed measure of exchangeability to evaluate a null hypothesis and two alternative hypotheses about the adaptiveness of the genetic code. The null hypothesis that there is no tendency for exchangeable amino acids to be assigned to similar codons can be excluded here as expected from earlier work. The first alternative hypothesis is that any such correlation between codon distance and amino acid distance is due to incremental mechanisms of code evolution, and not to adaptation to reduce deleterious effects of future mutations. More specifically, new codon assignments that occur by ambiguity reduction or by codon capture will tend to give rise to correlations, whether due to the condition of amino acid ambiguity, or to the condition of similarity between a new tRNA synthetase (or tRNA) and its parent. The second alternative hypothesis, the adaptive hypothesis, then may be defined as an excess relative to what may be expected given the incremental nature of evolution, reflecting true adaptation for robustness rather than an incidental effect. The results reported here indicate that most of the

nonrandomness in the amino acids to codon assignments can be explained by incremental code evolution, with a small residue of orderliness that may reflect code adaptation.

**Keywords** Genetic code · Mutations · Adaptiveness · Amino acids · Exchangeability

## Introduction

The idea that the genetic code is adaptive has a long history (Di Giulio 2005), despite disagreement as to what, precisely, it is adapted for—minimizing the effect of mutations, minimizing the effect of translation errors, or enhancing adaptive evolution. The adaptive hypothesis makes sense if, in some primordial era, many competing codes emerged, with the ultimate winner being a code with an unusually robust or effective organization (King 1971). Against the plausibility of this kind of scenario is the classic argument of Crick (1968) that any variability in the code will be costly once the number of genes grows large. By Crick's reasoning the code is largely a “frozen accident,” subject only to minor modifications. In fact, though the canonical genetic code is often mistakenly referred to as the “universal” code, it is not even constant on earth: various alternatives exist, typically differing only slightly from the canonical code (Knight et al. 2001).

Thus, a serious difficulty in studying code evolution is the lack of diverse codes for comparative analysis: the canonical genetic code appears to have arisen only once; it is clearly ancestral to all known cellular organisms, which have either the canonical code or a minor variant; and there is no evidence for the historical coexistence of alternative codes with major differences.

---

A. Stoltzfus (✉)  
Center for Advanced Research in Biotechnology,  
9600 Gudelsky Drive, Rockville, MD 20850, USA  
e-mail: arlin.stoltzfus@nist.gov

L. Y. Yampolsky  
Department of Biological Sciences, East Tennessee State  
University, Johnson City, TN 37614-1710, USA  
e-mail: yampolsk@etsu.edu

A second difficulty in evaluating hypotheses of code adaptation, all of which imply some congruence between the similarity of codons and the similarity of the encoded amino acids, is the lack of a measure of amino acid similarity that reflects the impact of amino acid changes on proteins but is not biased by the genetic code. That is, while the obvious measure to differentiate a pair of codons is some measure of genetic distance (e.g., number of differing nucleotides), it is not clear what is the best measure of the difference between two amino acids. The comment of Zuckerkandl and Pauling (1965) on the inadequacy of a priori views of amino acid similarity: “Apparently chemists and protein molecules do not share the same opinions regarding the definition of the most prominent properties of a residue”—is often repeated (e.g., Xia and Li 1998, p 33) Much of our understanding of the impact of amino acid changes on proteins comes from comparing naturally diverged proteins to yield a measure of evolutionary replaceability such as that reflected in a Dayhoff matrix. However, evolutionary replaceability reflects not just the exchangeability of amino acids in proteins, but codon similarity: the rate of a particular kind of evolutionary change depends on the rate of the underlying mutation, which in the case of amino acid changes depends on the genetic distance of the implicated codons (e.g., single nucleotide mutations are more likely than double or triple ones). For this reason, to use evolutionary replaceability as a proxy for amino acid exchangeability when evaluating genetic code evolution is to introduce logical circularity (Di Giulio 2001). Likewise, to prefer one biochemical index of similarity (e.g., hydrophobicity) over another because it fits evolutionary patterns, or to devise a compound index to fit evolutionary patterns (Grantham 1974), is to introduce a bias, given the enormous number of possible indices (Kawashima and Kanehisa 2000; Sneath 1966).

Recently, however, an unbiased measure of amino acid exchangeability in proteins, “EX” (Yampolsky and Stoltzfus 2005), was derived from a statistical meta-analysis of data from nearly 10,000 experimental amino acid exchanges. Since the exchanges are engineered under controlled conditions, EX is free of direct mutational effects of the genetic code. Importantly, EX values are asymmetric (i.e.,  $EX_{L \rightarrow F} \neq EX_{F \rightarrow L}$ ) and include not only the 150 “singlet” exchanges between amino acids whose codons are one nucleotide apart, but nearly all of the 202 doublet and 28 triplet exchanges that occur much less often in evolution (Whelan and Goldman 2004). Such a measure can be used as an objective means to evaluate whether there is any nonrandom tendency for similar codons to encode exchangeable amino acids.

A final difficulty is that even objective evidence of nonrandom organization of the genetic code would be open

to alternative interpretations that depend on uncertain models of evolutionary code change. Even if the actual path of code evolution has always been very narrow, without substantial competition between codes with different organizations, congruence between codon similarity and amino acid similarity might emerge as a side effect of incremental mechanisms of code evolution not guided by any global feature of code organization. This possibility, long considered in the literature of code evolution—where it is referred to obscurely as the effect of “history”—is an instance of a more general possibility (Wagner 2005) that “robustness” in evolved systems arises as a by-product of mechanisms of incremental change guided purely by local effects, rather than by selection for a global effect of robustness.

What incremental code changes may have happened, and how did they occur? A consensus chronology for the order of addition of different amino acids to the genetic code is provided by Trifonov (2004) based on various criteria judged to be relevant (e.g., evidence for synthesis under likely primordial conditions). Though the resulting consensus chronology is speculative, it is relatively robust to the precise choice of criteria (Trifonov 2004) and appears to be the best available basis for reconstructing incremental evolutionary changes to the code. Two mechanisms of such changes have been suggested: codon capture (Osawa 1995; Davis 1999) and ambiguity reduction (Davis 1999; Schultz and Yarus 1996). In the case of ambiguity reduction, amino acids  $i$  and  $j$  are partitioned to mutually exclusive subsets of a set of codons that previously were shared. Initially, amino acids  $i$  and  $j$  were ambiguously assigned, thus presumably highly exchangeable in proteins, and subsequently, the codons for  $i$  and  $j$  remain close in the genetic code (since they are partitions of the same original set). In the case of codon capture, some of the codons previously assigned to  $i$  are reassigned to  $j$ . In the classic scenario for codon capture, a codon for amino acid  $i$  is lost entirely from the genome before being reassigned to amino acid  $j$ , so that there is no necessary connection between previous sites of  $i$  and eventual sites of  $j$ . However, presumably those potential codon-capture pathways that may take place via a transitional period of ambiguity, in which  $i$  and  $j$  are assigned to some of the same codons, are more likely to take place because the intermediate shift in codon frequencies is less extreme.

Any evolutionary change in the genetic code implies a new specificity in the matching of codons and amino acids. This specificity is provided by tRNAs and the amino-acyl-tRNA synthetases (AARSs) that charge them with specific amino acids. Because an AARS recognizes both a tRNA (which specifies a codon) and an amino acid, incremental changes to AARSs will tend to assign similar amino acids

to similar codons (Cavalcanti et al. 2004; Salazar et al. 2003). The substrate specificity of AARS has long been considered as an important constraint on the evolution of the genetics code (Nagel and Doolittle 1995; Woese et al. 1966), but its effects on code organization have not been quantified. These effects could be addressed by using the “structural homology coefficient” ( $Q_H$ ) for AARSs computed from structural comparisons (O’Donoghue and Luthey-Schulten 2003).

From the above considerations, it is possible to define a set of hypotheses regarding the orderliness of the genetic code, and a set of tests to distinguish them. The null hypothesis is that there is no correlation between the similarity of amino acids (as measured by their exchangeability) and the similarity of their assigned codons (as measured by genetic distance). The first alternative hypothesis is that any such correlation is due to the incremental nature of code changes. The second alternative hypothesis, the adaptive hypothesis, is that there is an excess correlation beyond that expected merely from incremental mechanisms of evolution.

The alternative hypotheses may be distinguished by dividing amino acids into different categories of ordered pairs, according to the genetic distances between their codons and whether they are proposed to represent codon reassignments (via ambiguity reduction or codon capture). The classic adaptive interpretation of the genetic code is based on the observation that “singlet” pairs of amino acids, i.e., those whose codons differ by just one nucleotide, seem to be more similar than nonsinglet pairs. However, given that the proposed codon reassignments involve singlet pairs (Davis 1999; Trifonov 2000, 2004), the same inequality might be expected under the incremental hypothesis. The key is to distinguish the singlet pairs that are not proposed as codon-capture pairs. Under the incremental hypothesis, nonsinglets are indistinguishable from non-codon-capture singlets.

Evaluating this scheme yields the following results. First, the null hypothesis can be eliminated, as argued in many analyses done earlier by others (e.g., Freeland and Hurst 1998). Second, the observed orderliness of the genetic code can be explained largely by significantly higher than average conservativeness of codon-capture pairs, i.e., pairs of amino acids which consist of the presumed codon donor (“old” amino acid) and the codon recipient (“new” amino acid) added to the code by codon capture. A residual effect can be explained, under the incremental hypothesis, by AARS relationships. When these effects are allowed, there is no further signal representing specific evidence for the adaptive hypothesis. While this analysis is not definitive, it provides a useful framework for evaluating the apparent strength of competing claims about the causes of robustness.

## Materials and Methods

Exchangeability or EX (Yampolsky and Stoltzfus 2005) is defined such that  $EX_{ij}$  is the mean adjusted relative activity of a mutant protein that differs from the original by a single i-to-j exchange, based on some available set of such i-to-j exchanges, typically numbering several dozens. For the purpose of this analysis, instead of using the central tendency of effects in this set of i-to-j exchanges, we focus on the upper three-fourths of the distribution of effects, which corresponds roughly to a mutant protein activity of 75% or more of the wild type (the correspondence of the upper three-fourths with 75% activity is coincidental; see Eq. 1 and Fig. 2 of Yampolsky and Stoltzfus 2005). This measure is chosen to reflect the chance for an amino acid change to be merely tolerable (as opposed to benign). Such a measure would seem a natural choice given that the prevailing wisdom is that the genetic code is adaptive in some way that involves reducing the impact of errors (in replication, transcription, or translation) that might inactivate proteins. Indeed, of several matrices tested, the signal of apparent code adaptedness is strongest using  $EX_{75}$ . For instance, singlets differ more from nonsinglets when the difference is measured using  $EX_{75}$  as opposed to other versions of the EX matrix. Thus,  $EX_{75}$  is a conservative methodological choice in testing alternatives to adaptation. Nevertheless, the conclusions of this analysis do not depend qualitatively on the type of the EX matrix used.

For comparability purposes we also tested these hypotheses for a more traditional measure, used to estimate the severity of amino acid changes in the context of adaptedness of the genetics code—absolute difference in polar requirement,  $|\Delta PR|$  (as by Woese et al. 1966). This measure differs from the exchangeability by being symmetric, i.e.,  $|\Delta PR|_{i \rightarrow j} = |\Delta PR|_{j \rightarrow i}$ . The two measures are, not surprisingly, correlated (Spearman’s rank correlation,  $R = -0.43$ ,  $p < 0.0001$ ). Since both measures of effect used here are nonnormally distributed, nonparametric (Wilcoxon and Kruskal-Wallis) tests (SAS Institute 2002) were used.

Pairs of amino acids were classified into codon-capture and non-codon-capture pairs after Trifonov (2004) and (Davis 1999), assuming the following codon capture pairs (in this direction): V→L, L→I, L→F, R→Q, Q→H, I→M, S→R, N→K. Tyrosine (Y), tryptophan (W), and cysteine (C) are believed to be codon-capture amino acids as well, but they are hypothesized to have captured one of the initial stop codons. This means that, under the incremental hypothesis, expectations based on amino acid distances or tRNA distances do not apply. Therefore we ignore these amino acids.

Finally, note that a comparison of some severity-of-effect measure for singlet pairs of amino acids (amino acids with codons that can mutate into one another in a single

nucleotide substitution) to nonsinglet pairs, is comparable but not precisely equivalent to the Monte Carlo approach estimating error-proneness of simulated codes with only single nucleotide changes assumed possible (Freeland and Hurst 1998; Freeland et al. 2003).

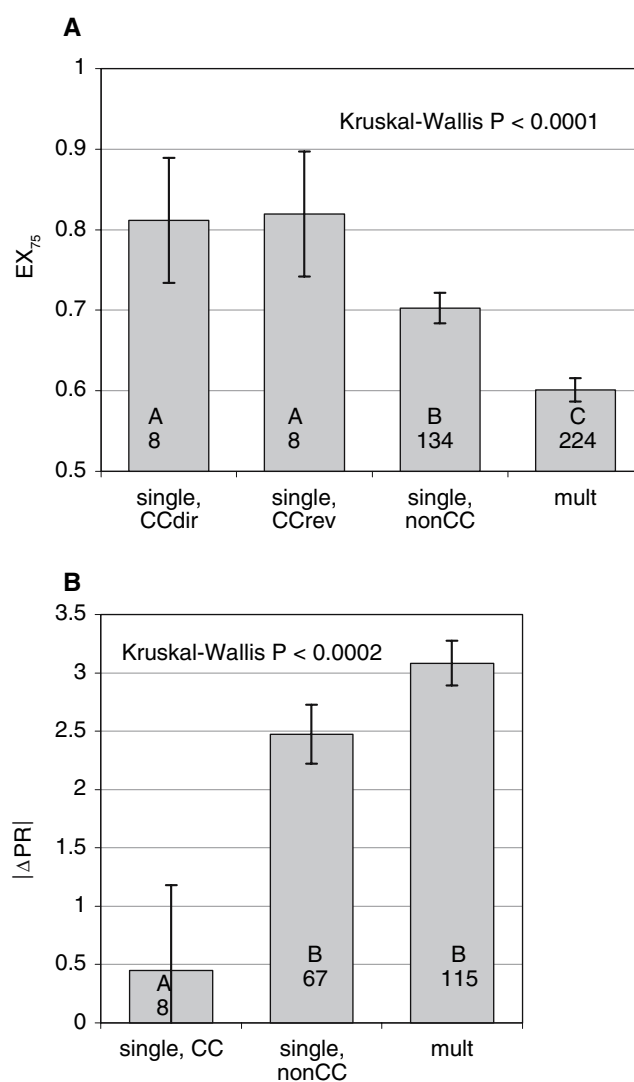
## Results and Discussion

As a measure of amino acid similarity suitable for the present purposes, we use  $EX_{75}$ , a version of EX that reflects, not the mean effect of exchanging  $i$  and  $j$ , but the chance of a tolerable effect, where “tolerable” excludes only the lowest quartile of outcomes (see Materials and Methods). For comparison with some earlier studies, we also use the “polar requirement” metric  $|\Delta PR|$  (see Materials and Methods). Mean values of  $EX_{75}$  and  $|\Delta PR|$  calculated for singlet pairs (both codon-capture and non-codon-capture) and nonsinglet amino acid pairs are shown in Fig. 1.  $EX_{75}$  is higher for the codon-capture pairs, regardless of the direction, than for non-codon-capture singlets (Wilcoxon’s test,  $p < 0.01$ ) and higher for non-codon-capture singlets than for nonsinglets (Wilcoxon’s test,  $p < 0.002$ ). Overall differences are significant (Kruskal-Wallis test) at  $p < 0.0001$ . A similar pattern is observed for  $|\Delta PR|$ .

As a first approximation, these results lend support both to the incremental hypothesis and to the adaptive hypothesis, since codon-capture pairs have the highest exchangeability (supporting the incremental hypothesis) but non-codon-capture singlets still have higher exchangeability than nonsinglets (supporting the adaptive hypothesis).

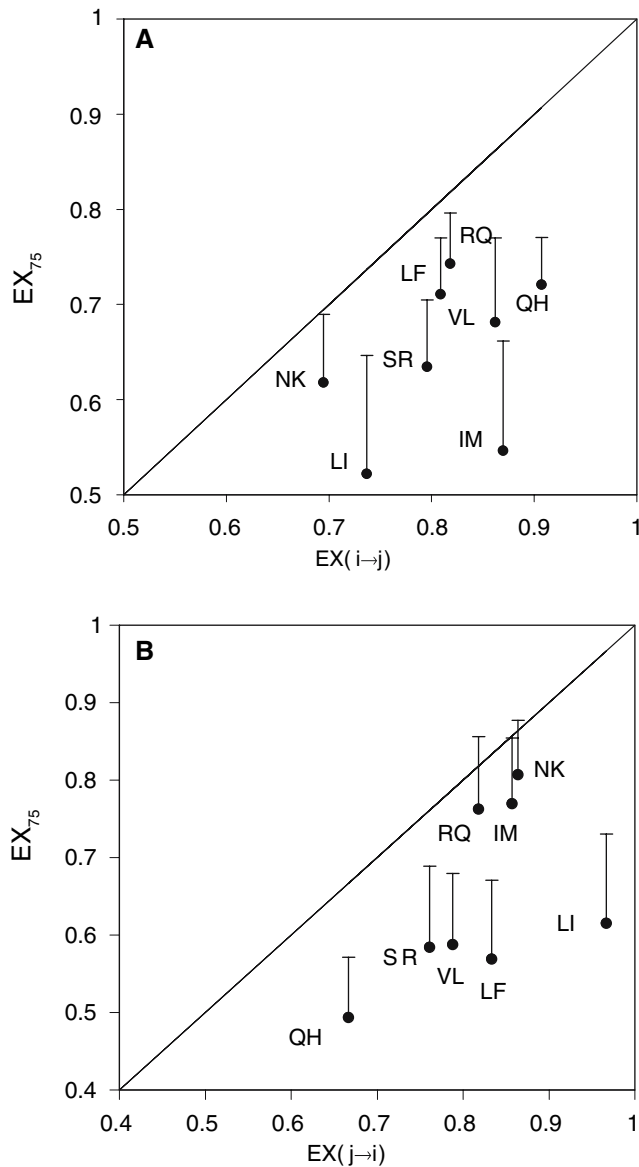
To further characterize the unusual nature of the hypothesized codon-capture pairs, we considered whether the donor amino acids (from which codons are unassigned) were the most likely donors for the recipient amino acid (to which codons are assigned), among all amino acids present in the code at the time of this addition (assuming the “consensus” order of amino acid addition (Trifonov 2004)). The results of this comparison are shown in Fig. 2A. In all eight cases, the codon donor is significantly more exchangeable by the codon recipient than the average available donor. As shown in Fig. 2B, this is also in five of eight cases for the reverse comparison: the recipient (amino acid to which codons are assigned) is significantly more easily substituted by the donor than by an average amino acid already present in the code. In the other three cases (Q, M, and K), the presumed codon donor is merely average in exchangeability.

To test the implications of incremental change in AARS substrate specificity, we correlated the mean  $EX_{75}$  for pairs of amino acids with the structural relationships of their respective AARSs (O’Donoghue and Luthey-Schulten 2003). This is done separately for each of the two AARS



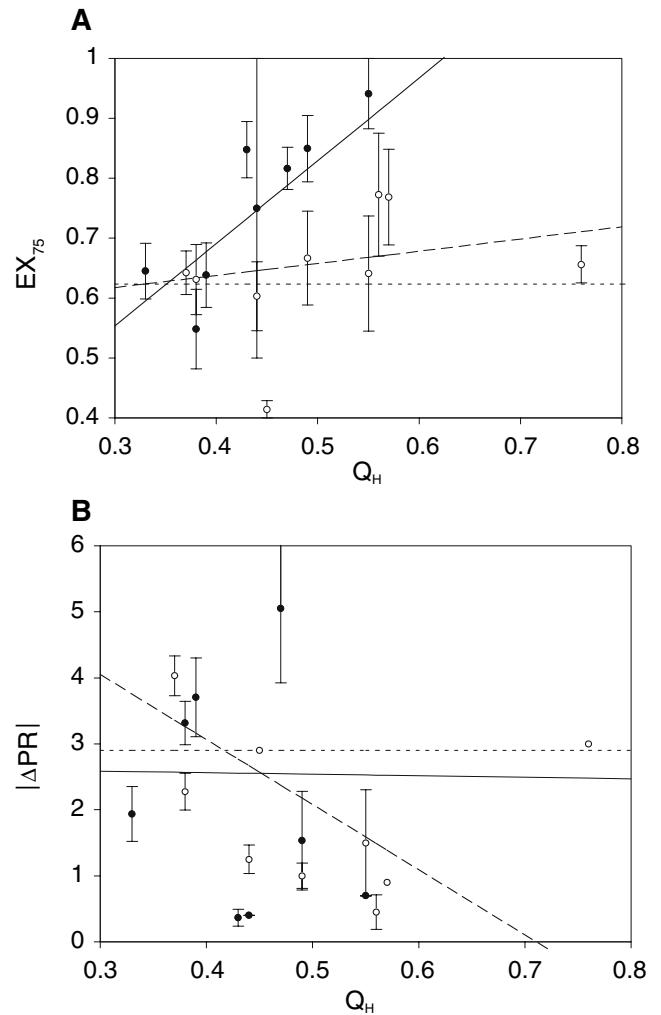
**Fig. 1** Amino acid difference measures for categories of amino acid pairs distinguished with respect to hypotheses of interest. **A** Mean (and standard error) amino acid exchangeability ( $EX_{75}$ ). **B** Mean absolute difference (and standard error) in polar requirement,  $|\Delta PR|$  (as in Woese et al. 1966). The categories distinguish codon-capture pairs (“single, CC dir”), reversed codon-capture pairs (“single, CC rev”), singlet non-codon-capture pairs (“non-CC”), and pairs where the codons differ by multiple nucleotides (“mult”). Directional CC comparisons are not possible for  $|\Delta PR|$ , which is symmetric. Each bar has a number indicating the number of data and a letter indicating which columns differ significantly by Dunnett’s test for multiple comparisons (with “mult” as the control group [SAS Institute 2002])

classes (omitting pairs of amino acids involving alanine, since the structure of alanyl-tRNA synthetase is not available). In corroboration of a similar result using a different approach (sequence-based phylogenies and a combination of correlated amino acid properties (Cavalcanti et al. 2004)), we observe a positive correlation between mean  $EX_{75}$  and  $Q_H$ , the measure of structural homology between AARSs (Fig. 3A). As expected,  $EX_{75}$  is higher for same-class pairs than across-class pairs (dotted



**Fig. 2** Putative codon-capture pairs show unusual amino acid exchangeability relative to hypothetical alternatives. In **A** and **B**, each data point labeled with two letters,  $X_iX_j$ , refers to the putative codon-capture event in which codons for  $X_i$  were reassigned to (captured by) amino acid  $X_j$ . In **A**, the mean  $EX_{75}$  value (with upper tail of 95% CI) for all other possible donor amino acids (i.e., all other amino acids presumed to be present in the code at the time of addition of  $j$ ) is shown as a function of  $EX_{75}(i \rightarrow j)$ , that is, the exchangeability value for the putative codon-capture pair. In **B**, the mean  $EX_{75}$  from all other possible donors is shown as a function of  $EX_{75}(j \rightarrow i)$ , i.e., exchangeability in the reverse direction. Codon capture pairs and consensus order of amino acid additions follow Trifonov (2004)

lines in Fig. 3) and this difference increases with structural similarity of the AARSs. This correlation, however, is significant only for amino acids served by AARSs in Class I (Spearman's rank correlation,  $\rho = 0.31$ ,  $p < 0.003$ ) and not in Class II ( $\rho = 0.009$ ,  $p > 0.9$ ). The correlation



**Fig. 3** Relationship of  $EX_{75}$  (**A**) and absolute difference in polar requirement (**B**) to structure homology between corresponding AARSs. Absolute difference in polar requirement,  $|\Delta PR|$ , is computed according to Woese et al. (1966), and structural homology coefficient,  $Q_H$ , according to O'Donoghue and Luthey-Schulten (2003). Filled circles, solid regression lines—AARSs Class I (Fig. 11 of O'Donoghue and Luthey-Schulten 2003); open circles, dashed regression lines—AARSs Class II (Fig. 12 of O'Donoghue and Luthey-Schulten 2003). Pairs of amino acids including Ala are omitted due to lack of structure data for the alanyl-tRNA synthetase. Dotted lines: mean  $EX_{75}$  and mean  $\Delta PR$  for pairs of amino acids served by AARSs from different classes

between  $Q_H$  and  $|\Delta PR|$  reveals a difference: the negative correlation expected under the incremental hypothesis (negative because  $Q_H$  is a measure of similarity while  $|\Delta PR|$  is a measure of difference) is significant for synthetases of class II ( $\rho = -0.68$ ,  $p < 0.0001$ ) but not class I ( $\rho = 0.075$ ,  $p > 0.60$ ). At the risk of overinterpretation, this suggests that polarity of the substrate amino acid constituted a more important incremental constraint in the evolution of new class II synthetases, while overall exchangeability was more important in the case of class I synthetases.

**Table 1** Comparison of  $EX_{75}$  values for categories of amino acid pairs distinguished with respect to hypotheses of interest

Genetic code affinity	Different classes			Same class			Same class, residuals of $EX_{75} \sim Q_H$		
	<i>N</i>	Mean	SE	<i>N</i>	Mean	SE	<i>N</i>	Mean	SE
Single CC	6	0.80	0.09	10	0.83	0.07	10	0.10	0.07
Non-CC	70	0.71	0.03	64	0.70	0.03	54	0.00	0.03
Multiple	121	0.56	0.02	103	0.64	0.02	95	-0.01	0.02
	$p < 0.0001$			$p < 0.018$			$p > 0.27$		

Note. *N*, number of values; SE, standard error; *p*, probability from Kruskal-Wallis test for difference between genetic code affinity classes; CC, codon capture

In both cases, assuming that the  $EX_{75}$  and  $|\Delta PR|$  also can serve as a measure of amino acid exchangeability as AARS substrates, newly evolving AARSs appear to retain a higher degree of conservation of their substrate specificity. It is natural to assume that structurally and phylogenetically close AARSs also retained specificity to similar tRNAs. Such conservation contributes to the pattern of similar amino acids being assigned to similar codons and can therefore be interpreted erroneously as the signature of adaptedness of the genetic code. To test this hypothesis we calculated mean  $EX_{75}$  values as in Fig. 1, but for pairs of amino acids served by AARSs from the same class and from difference classes separately (Table 1). Note that in Table 1, the mean and standard deviation are shown for the amino acid pairs in the relevant category, where the rows are defined by the distinction of single- vs. multiple-nucleotide changes, and by the distinction of codon-capture (“CC”) vs. non-codon-capture (“non-CC”) pairs, and the columns are defined by the distinction of amino acids served by amino-acyl tRNA synthetases of the same class vs. those served by different classes.

For the pairs of amino acids served by AARSs from the same classes we also calculated the differences between residual values of linear regression of  $EX_{75}$  on  $Q_H$ , estimated for each of the two AARS classes separately (rightmost three columns in Table 1). As Table 1 shows, the difference in  $EX_{75}$  between pairs of amino acids with different genetic code affinities disappears when structural homology of the AARSs is accounted for. The results are similar using  $|\Delta PR|$  instead of  $EX_{75}$  (not shown).

Thus, the apparent higher-than-average exchangeability of singlets can be explained entirely by structural similarity of AARSs serving amino acids assigned to similar codons. On the other hand, this only explains the differences between singlets and nonsinglets among pairs of amino acids served by AARSs from the same classes. This difference remains statistically significant for the pairs of amino acids served by AARSs from different classes, to which this same reasoning of evolutionary conserved

structure does not apply, a result that would seem to support the adaptive hypothesis.

In summary, we evaluated the possibility that incremental mechanisms of evolution might explain orderliness of the genetic code, rather than adaptation for robustness. This analysis focused on two aspects of incremental change that might explain a tendency for amino acids with similar codons to be similar in their properties: the possibility of transitional ambiguity favoring the exchangeability of amino acids involved in specific codon-capture events and the conservativeness of evolutionary changes in AARSs (their amino acid and tRNA specificities) during early evolution. We found evidence of both factors in the structure of the canonical genetic code. Together these two factors account for a substantial portion, but not all, of the congruence between similarity of codons and similarity of corresponding amino acids. These two factors do not account for the excess exchangeability of non-codon-capture singlet pairs served by AARSs from different classes, a result that supports the adaptive hypothesis.

Two important caveats should be attached to the arguments presented here. First, this analysis depends on highly speculative conjectures about the order of ancient events, namely, the addition of new amino acid-codon pairs to the genetic code. Though these conjectures are based largely on the interpretation of biochemical synthesis pathways for amino acids, this reasoning is not entirely independent of the kinds of data examined here. Second, although we introduce exchangeability and AARS structural similarity as factors to account for the kinds of orderliness that may arise by incremental mechanisms, we do not present a quantitative model that predicts the exact form of the orderliness. This is important because the predictive factors for orderliness due to adaptation would be similar.

Finally, it should be noted that the question of whether the “robustness” of genetic encoding is adaptive in origin, or is a by-product of incremental mechanisms of code evolution, does not alter the proximate implications of this “robustness” for predicting the resilience of gene expression in the face of errors in replication, transcription, or translation.

**Acknowledgments** The authors thank Steve Freeland and three anonymous reviewers for advice and comments. The identification of specific commercial software products in this paper is for the purpose of specifying a protocol, and does not imply a recommendation or endorsement by the National Institute of Standards and Technology.

## References

- Cavalcanti AR, Leite ES, Neto BB, Ferreira R (2004) On the classes of aminoacyl-tRNA synthetases, amino acids and the genetic code. *Orig Life Evol Biosph* 34:407–420
- Crick FHC (1968) The origin of the genetic code. *J Mol Biol* 38:367–379
- Davis BK (1999) Evolution of the genetic code. *Prog Biophys Mol Biol* 72:157–243
- Di Giulio M (2001) The origin of the genetic code cannot be studied using measurements based on the PAM matrix because this matrix reflects the code itself, making any such analyses tautologous. *J Theor Biol* 208:141–144
- Di Giulio M (2005) The origin of the genetic code: theories and their relationships, a review. *Biosystems* 80:175–184
- Freeland SJ, Hurst LD (1998) The genetic code is one in a million. *J Mol Evol* 47:238–248
- Freeland SJ, Wu T, Keulmann N (2003) The case for an error minimizing standard genetic code. *Orig Life Evol Biosph* 33:457–477
- Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185:862–864
- Kawashima S, Kanehisa M (2000) AAindex: amino acid index database. *Nucleic Acids Res* 28:374
- King JL (1971) The influence of the genetic code on protein evolution. In: Schoffeniels E (ed) *Biochemical evolution and the origin of life*. North-Holland, Viers, pp 3–13
- Knight RD, Freeland SJ, Landweber LF (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* 2:49–58
- Nagel GM, Doolittle RF (1995) Phylogenetic analysis of the aminoacyl-tRNA synthetases. *J Mol Evol* 40:487–498
- O'Donoghue P, Luthey-Schulten Z (2003) On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol Mol Biol Rev* 67:550–573
- Osawa S (1995) *Evolution of the genetic code*. Oxford University Press, New York
- Salazar JC, Ahel I, Orellana O, Tumbula-Hansen D, Krieger R, Daniels L, Soll D (2003) Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates. *Proc Natl Acad Sci USA* 100:13863–13868
- SAS Institute (2002) *JMP statistics and graphics guide*. SAS Institute, Cary, NC
- Schultz DW, Yarus M (1996) On malleability in the genetic code. *J Mol Evol* 42:597–601
- Sneath PHA (1966) Relations between chemical structures and biological activity in peptides. *J Theoret Biol* 12:157
- Trifonov EN (2000) Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 261:139–151
- Trifonov EN (2004) The triplet code from first principles. *J Biomol Struct Dyn* 22:1–11
- Wagner A (2005) *Robustness and evolvability in living systems*. Princeton University Press, Princeton, NJ
- Whelan S, Goldman N (2004) Estimating the frequency of events that cause multiple-nucleotide changes. *Genetics* 167:2027–2043
- Woese CR, Dugre DH, Dugre SA, Kondo M, Saxinger WC (1966) On the fundamental nature and evolution of the genetic code. *Cold Spring Harb Symp Quant Biol* 31:723–736
- Xia X, Li WH (1998) What amino acid properties affect protein evolution? *J Mol Evol* 47:557–564
- Yampolsky LY, Stoltzfus A (2005) The exchangeability of amino acids in proteins. *Genetics* 170:1459–1472
- Zuckerklund E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) *Evolving genes and proteins*. Academic Press, New York