

# The Genetic Code—More Than Just a Table

D. Berleant<sup>1,5</sup>, M. White<sup>2</sup>, E. Pierce<sup>1</sup>, E. Tudoreanu<sup>1</sup>, A. Boeszoermyi<sup>3</sup>,  
Y. Shtridelman<sup>4</sup> and J. C. Macosko<sup>4,5</sup>

<sup>1</sup>Department of Information Science, University of Arkansas at Little Rock, 2801 S. University Ave., Little Rock, AR 72204, USA

<sup>2</sup>Rafiki Inc., Bloomington, IN, USA

<sup>3</sup>Karl-Franzens-Universität Graz, Graz, Austria

<sup>4</sup>Department of Physics, Wake Forest University, 1834 Wake Forest Rd., Winston-Salem, NC 27109, USA

<sup>5</sup>Emails: berleant@gmail.com; macoskjc@wfu.edu

**Abstract** The standard codon table is a primary tool for basic understanding of molecular biology. In the minds of many, the table's orderly arrangement of bases and amino acids is synonymous with the true genetic code, i.e., the biological coding principle itself. However, developments in the field reveal a much more complex and interesting picture. In this article, we review the traditional codon table and its limitations in light of the true complexity of the genetic code. We suggest the codon table be brought up to date and, as a step, we present a novel superposition of the BLOSUM62 matrix and an allowed point mutation matrix. This superposition depicts an important aspect of the true genetic code—its ability to tolerate mutations and mistranslations.

**Keywords** Genetic code - Codon table - Biological pedagogy - Code optimization - Point mutation - BLOSUM62 - PAM250

---

## Introduction

What is the *genetic code*? Metaphors that have been proposed all have significant limitations [46]. Analysis of the relevant hits in the first three pages that Google returned for the query “genetic code” (see Supplemental Materials) reveals how the concept is typically understood.

As envisioned by the codon table (19 citations)

Instructions for protein synthesis (3 citations)

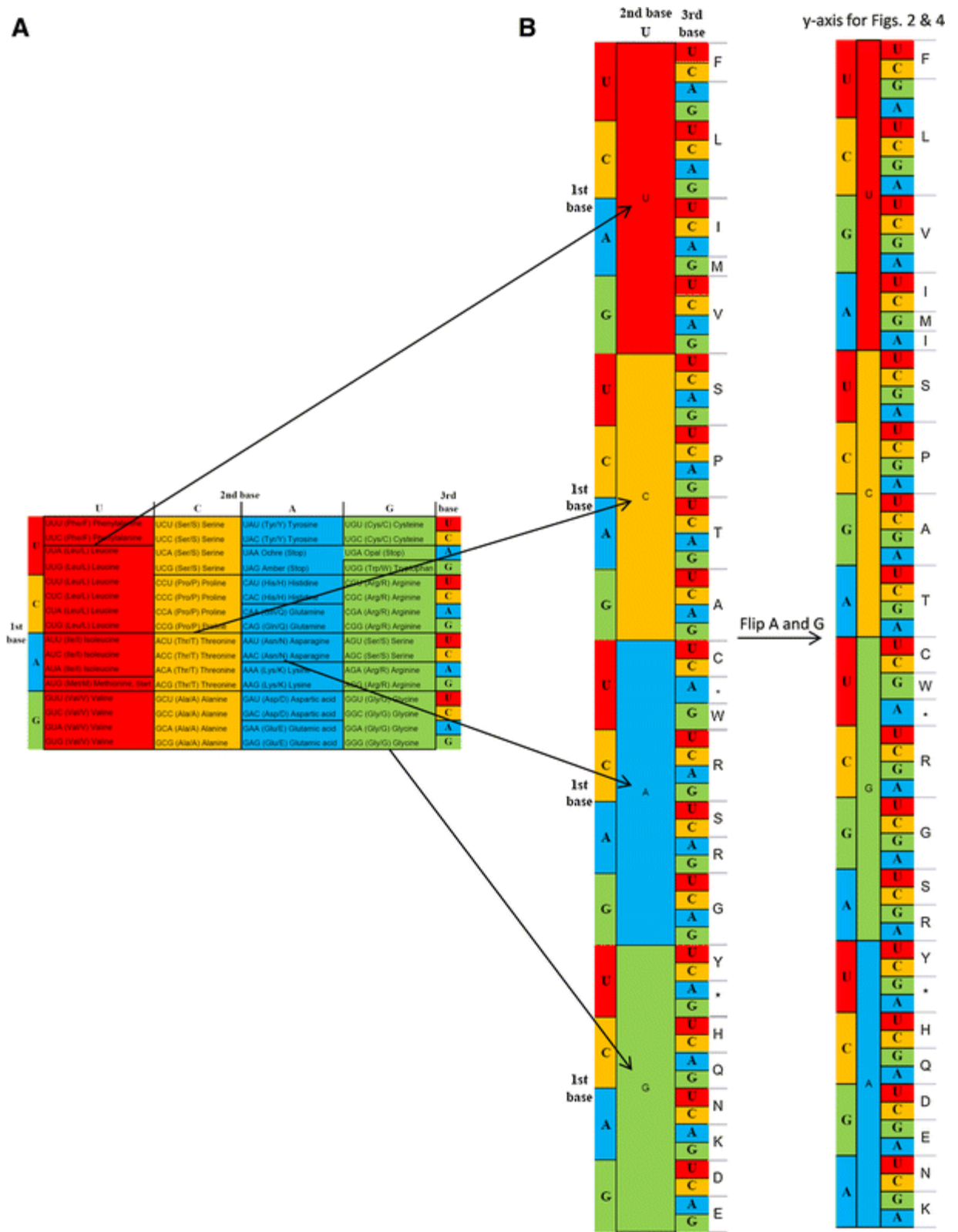
The mechanism for storing genetic information (2 citations)

DNA sequences (2 citations)

Other (1 citation)

Clearly there is a common perception that the standard codon table (Fig. 1a; [8]) is synonymous with, and adequately represents, the genetic code. Indeed, it was discovered early on that a near universal, near perfect correlation exists between nucleic acid triplets (the codons) and amino acids in living things [32, 36, 37]. This correlation is succinctly embodied in the standard codon table. However, caution is warranted because this table represents only a modest subset of the complexity of translation implied by an organism's DNA, and even less when transcription and other molecular relationships are considered. A more richly explanatory representation would better capture the genetic code concept.

The purpose of this article is threefold. Firstly, we intend to communicate technical details and references, to a proficient audience, that support a richer and more sophisticated view of the genetic code than is often held by other audiences. Secondly, we hope to increase awareness in such readers about the impoverished, even distorted, understanding of the crucial genetic code concept among many not at the forefront of the field. Meeting these first two goals may help the reader to transmit a deeper appreciation for the genetic code when teaching, communicating with science writers, authoring semi-popular materials, and engaging in other outreach activities. The third goal is to illustrate the continuing potential for new and improved representations with an example. Meeting this goal may help inspire others to work toward other representations that capture, as well as possible, the full beauty and grandeur of the genetic code.



**Fig. 1** The codon table in its standard form (**a**), which is considered by many as *the* “genetic code.” Ordering is UCAG, top to bottom, left to right [8]. The first codon position

specifies the row, the second specifies the column, and the third specifies the order within a given row. The color scheme helps guide the eye as columns are stacked and reshuffled to produce the axes in Figs. 2 and 4 **(b)**. The motivation for reshuffling by flipping the positions of A and G is to preserve the general trend of decreasing amino acid hydrophobicity from column U through columns C and G and finally to column A, whereas the motivation for the standard UCAG ordering is to maximize codon groupings, (e.g., all the isoleucine codons are grouped together in the standard ordering)

---

## Limitations of the Standard Codon Table

The traditional codon table fails to capture current understanding along a number of dimensions. As examples, (a) there are many exceptions to the standard code, (b) “synonymous” codons are not always synonymous, (c) there are layers of information overlapping mere amino acid specification, and (d) the codon table is highly optimized to reduce the consequences of mutation and mis-translation. These dimensions are reviewed next.

### There are Exceptions to the Standard Code, thus Ambiguities in the Standard Table

The standard codon table implies that each codon has a single interpretation. But this does not always hold. Three types of such exceptions are given below, for which ambiguities are resolved by molecular context, environmental context, or cellular context. For those resolved by molecular context we give three examples.

(1a) *Molecular context—selenocysteine*. The unusual amino acid selenocysteine, abbreviated Sec and U [48], is a modification of cysteine in which the sulfur atom is replaced by a selenium atom, selenium being directly below sulfur in the periodic table. The RNA codon for selenocysteine is UGA [10]—though in the standard codon table UGA is a stop codon. In fact, whether UGA codes for stop or selenocysteine is context dependent.

(1b) *Molecular context—pyrrolysine*. The unusual amino acid pyrrolysine (abbreviated Pyl and O), if present, uses the codon UAG. Like UGA, UAG is listed as a termination signal in the standard codon table. To determine whether a Sec or Pyl amino acid is inserted into a growing polypeptide chain, or whether the chain undergoes termination instead, the translation machinery uses molecular context such as structures in the 3' UTR (untranslated region) of the mRNA or adjacent nucleotides that redefine the stop codon [23]. Thus, the UGA and UAG codons, without the appropriate context, are ambiguous.

(1c) *Molecular context—methionine*. A further example, and the most familiar, is the methionine codon, AUG, which also codes for start in some situations though not in others.

These three molecular context examples are deterministic in nature (though ambiguous by themselves, their molecular context determines their interpretation). However, for the environmental and cellular context cases described below, the codon's surrounding nucleotides do not disambiguate.

(2) *Environmental context.* The “CUG ambiguity” [41] was discovered in some species of the yeast genus *Candida*, an organism known for causing illnesses in humans and animals. In these organisms, the CUG codon sometimes translates as serine and sometimes as leucine. Interestingly, experimental evidence shows that this ambiguity can be functionally useful, leading to improved stress response under certain environmental conditions [42]. Another example of this type of ambiguity is the inability of animals to distinguish between methionine and the rare amino acid selenomethionine in translation, so that selenomethionines can be unpredictably incorporated into proteins where ordinary methionine would normally be [44].

(3) *Cellular context.* In a third type of codon ambiguity, a codon may translate differently depending on sub-cellular location, such as the codon assignment variation in mitochondria (e.g., [50]). The fourth type of codon ambiguity is perhaps the best known, and occurs when a codon is interpreted one way in one species and another way in a different species. One notable case, the AGG codon, depending on the species translates to arginine, serine, or glycine, is a stop codon, or is simply unassigned (e.g., [27]) with no tRNA present to decode it ([42], Table 3). This type of ambiguity is becoming engineerable [5], and intentionally produced strains of organisms with modified translational codes are expected to have significant applications.

Fortunately, the inability of the classical codon table to express codon ambiguity is readily fixed. Entries in the classical form of the table can and have been augmented to contain not just one interpretation as is usually the case, but rather each of the various amino acid and other codes found to be associated with any given nucleotide triplet. Annotations about *when* each interpretation applies would be simple to add as well, but potentially cumbersome depending on the quantity of annotation information. Yet even thus augmented, the codon table still has serious additional limitations. As we shall see in the following sections, these limitations include failures to: (i) identify the functional differences in “synonymous” codons, (ii) reveal the layers of information in addition to mere amino acid specification, and (iii) expose the resulting optimized robustness of the standard codon-to-amino acid mapping function.

## **“Synonymous” Codons Are Not Necessarily Equivalent**

The codon table lists many cases of different codons that code for the same amino acid. In fact, the standard table maps 64 codons to 21 interpretations (20 amino acids and the stop codon), an average of about three codons per interpretation;

only two of the 21 are coded by just one codon (these are methionine and tryptophan). But the different codons to which the table attributes the same interpretation are not, in fact, equivalent. There are four main types of non-equivalence of apparently synonymous codons.

(1) *Different species possess synonymous codons in different proportions.* These patterns of “codon usage bias” could not arise by purely random variation, implying biological causation of such bias. The details of this preference for some synonymous codons over others has been linked to a variety of factors—gene expression level, gene translation initiation signal, protein structure, mutation frequency and patterns, etc. [1]. Bacterial fitness in particular tends to rely on efficiency, and having fewer commonly used codons increases potential efficiency of translation by reducing the tRNA concentration required for fast translation. On the other hand, codon synonymity can enhance robustness to mutation by permitting protein sequences to be conserved despite changes in the DNA.

The degree of codon usage bias can be dramatic. CGC is the RNA codon for arginine 33% of the time in fruit flies (*Drosophila melanogaster*) according to figures provided by the Kazusa DNA Research Institute, but only 18% in humans [35]. On the other hand, AGG was the codon for arginine 21% of the time in humans, but only 11% of the time in *D. melanogaster*, and just 7% of the time in *E. coli*. These figures can vary considerably depending on the cited source due to variations in counting technique, but that there are major differences across species is well-known (Rob Knight, personal communication, January 5, 2009).

In a classic example of how these biases enter the real world of genetic engineering, the codons for GFP (green fluorescent protein), which is derived from the *Aequorea victoria* jellyfish, required adjustment to match the usage biases of a new host into which the gene was introduced before high expression levels of this important fluorescent marker were achieved [21].

(2) *Different codons for the same amino acid can require different amounts of time to be translated into the amino acid.* Nakamura and Sugiura [34] demonstrated this in tobacco chloroplasts and showed that it is not necessarily correlated to codon usage. Translation time can be functionally important because it affects a protein’s availability in an organism. Moreover, translational pauses can affect protein folding, as discussed below. At least one patent application has been based on the differential translation efficiencies of synonymous codons [12].

(3) *Different codons for the same amino acid can have different effects on the folding of an RNA molecule.* Different RNA foldings can have major functional implications. For example, whether a UGA triplet codes for STOP or for the unusual amino acid selenocysteine depends on other, nearby cues that affect folding. Often, only a specific nearby stem–loop structure will lead to the creation of a selenoprotein [20]. In *E. coli* a 17-nucleotide sequence, nominally

GGUUGCAGGUCUGCACC [29] but with numerous possible variations [40], beginning 11 nucleotides downstream from the UGA triplet is part of the code for selenocysteine. According to Sandman et al. [40], “It may be possible to [design] a SECIS [selenocysteine insertion sequences] that allows expression of the native downstream amino acid sequence of the protein.” In other words, synonymous codons could be judiciously chosen so as to maintain the “native downstream amino acid sequence” while still achieving the proper stem–loop structure in the mRNA that triggers selenocysteine insertion [40]. More recently, it was found that cues in the 3′ untranslated region of a certain ciliate (*Euplotes crassus*) can cause the UGA (stop) codon to code either for cysteine or for selenocysteine [47].

(4) *Different codons for the same amino acid can have different effects on the folding of proteins during translation.* In the *E. coli* example above, a turn which connects two  $\alpha$ -helices in a protein called *Echinococcus granulosus* Fatty Acid Binding Protein1 (EgFABP1) was tested in vivo to determine the effect of mutations substituting synonymous codons for various amino acids in the turn [7]. It was found that the fold of the resulting protein changed significantly for one of the seven sets of synonymous substitutions tested. The fold difference was detected by the functional changes it produced in the resulting protein as well as by a reporter gene sensitive to the misfold. Differential folding due to these so-called silent mutations appears to be connected to their differential use of alternative tRNA molecules to carry the same amino acid building block during protein construction [39]. Their results bear on the problem of prion proteins, the misfolding of which causes mad cow disease, scrapie in sheep, and Creutzfeldt–Jakob disease in humans.

## Mere Amino Acid Specification is Not Enough

Improving the informational richness of the codon table begins by recognizing that the codon table in its typical forms (Fig. 1a) lists only the names of the amino acids. This is readily achieved.

## Enriching Information Content

Like the periodic table which lists, in addition to element names, many of their properties, the codon table would be more useful if it listed the amino acids’ physicochemical properties such as hydrophobicity, electric charge, and size [33]. A simple solution sometimes used is to provide a ball-and-stick representation of each side chain adjacent to each amino acid name [38].

A less obvious but equally important type of information that is missing relates to the mapping of tRNA molecules, which carry amino acids to the ribosome for incorporation into a growing protein. The tRNA molecules contain anticodon triplets that bind to complementary triplets in the mRNA, but while the first two nucleotides pair in the ordinary Watson–Crick way, the third pairs using so-called wobble rules, which are different. This helps explain such phenomena as the

redundancies in the triplet codes that make up the codon table. In addition, tRNAs with the same anticodons can exist in multiple isodecoder forms [19]. Since translation is mediated by tRNA molecules, which map to the mRNA codons they bind to and to the amino acids they carry to the growing protein in interesting ways, it would be useful if wobble rules and tRNA species were expressed in a codon table representation.

Finally, other important properties of genetic coding are also ignored by the standard codon table. These include that: (i) similar codons tend to code for physico-chemically similar amino acids, (ii) if position 2 of a codon is U, the codon is for a hydrophobic amino acid, (iii) heavier amino acids tend to have fewer codons, and (iv) frequently appearing amino acids tend to have more codons [28]. Using the periodic table as a parallel, it would enhance the value of the codon table if, for example, such similarities among amino acids could be presented as groupings within it.

## Protein Folding

In discussing additional layers of information, we have so far focused (like the codon table itself) on individual amino acids. In fact, the key property of a protein is very often not its specific amino acids, but the physical structure of its fold. The essential building blocks of the fold are the dihedral angles between pairs of amino acids [43]. These angles do not exist as properties of single amino acid molecules (and therefore of nucleotide triplets), but rather are a property in large part of pairs of amino acids (and hence of nucleotide sextuplets). The sequence of each pair's three backbone dihedral angles ( $\phi$ ,  $\psi$ , and  $\omega$ ) along with each side chain's  $\geq 5$  dihedral angles ( $\chi_1$ – $\chi_5$ ) uniquely determines the physical shape and functional abilities of the protein [6]. Thus, the genetic code is not just how a string of DNA letters is converted, via mRNA, into amino acid letters, but how the DNA information becomes the three-dimensional information of a sequence of dihedral angles. The codon table is only a part of this process—a distinction that should be made explicit in presenting the “genetic code.”

## Symmetries

Some common mutations in DNA sequences, such as frame shifts, complements, inversions, and combinations thereof, lead to wholesale changes in all of the codons in transformed sequences. However, all of the changed codons are changed by the same kind of transformation. A deeper understanding of underlying symmetries in the genetic code would bring to light the possibility that the resulting products of translation after such transformations—typically thought of as merely randomized—can have higher evolutionary value than truly randomized sequences [49].



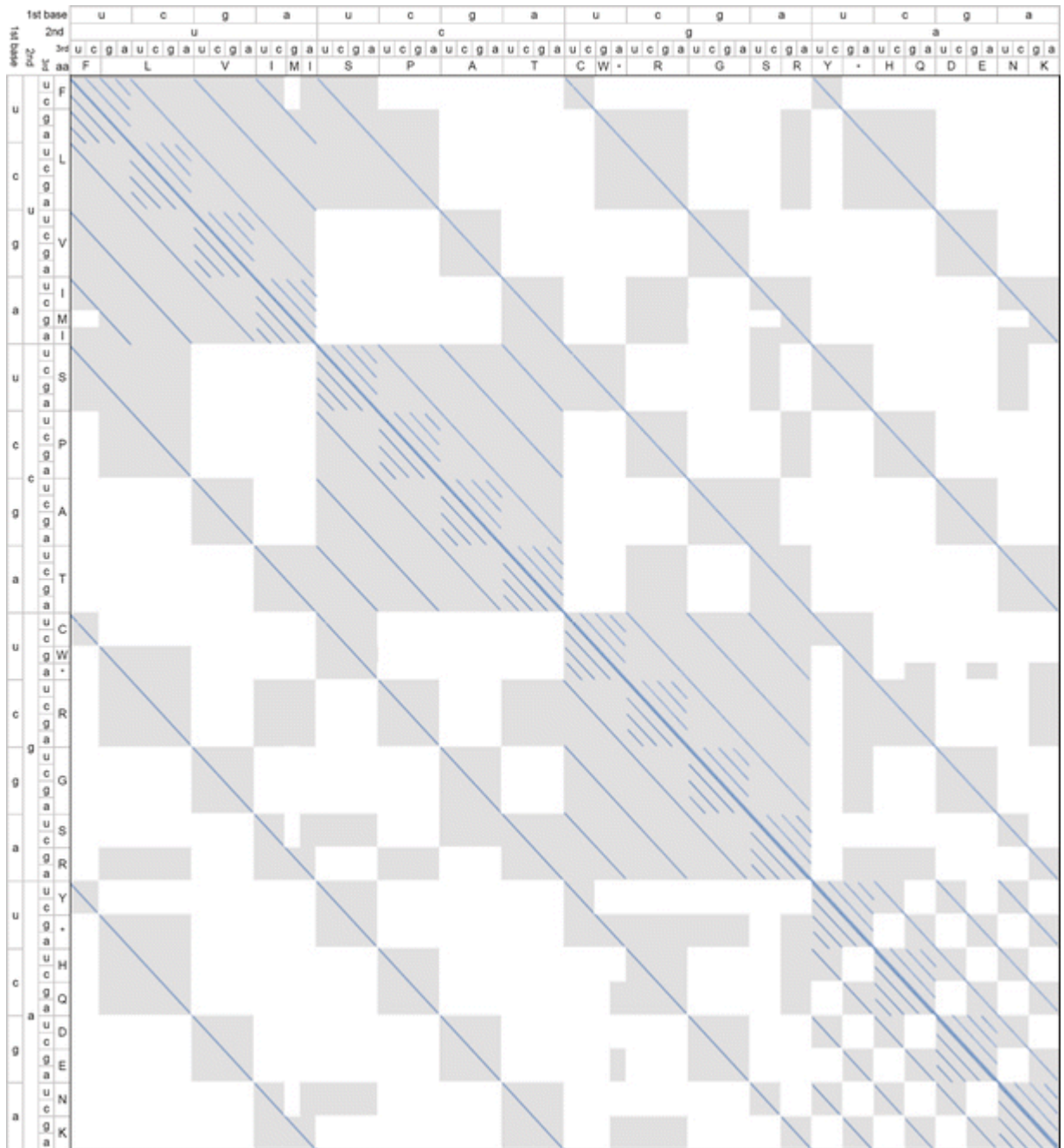
## The Codon Table Buffers Mutations and Mistranslations

Soon after the discovery of the coding relationship between nucleotide bases and amino acids, the codon table was described by Francis Crick [8] as a “frozen accident.” Yet in recent years this coding relationship has been shown to be optimal in many respects (e.g., [4]). The most undisputed aspect of code optimality is its robustness with respect to point mutations and mistranslations [15, 28], also called the “load minimization” or “error-buffering capacity” of the genetic code. Unfortunately, the standard codon table does not reveal this important aspect of the genetic code.

To see this in slightly more detail, note that some of the 64 codons can interconvert by single point mutations. A point mutation matrix, the axes of which can be constructed in many different orderly ways (e.g., Fig. 1b), indicates whether a given interconversion is allowed by a single point mutation (Fig. 2). With single point mutations, only 9 out of 63 of the interconversions are possible starting from any given codon. In other words, a codon can convert to any of the 63 other codons by 1, 2, or 3 point mutations but can convert to only 9 other codons by 1 point mutation. For example, AAA can convert to GAA but not to GGA (nor GGG) via a single point mutation.

As a consequence of the codon-to-amino acid mapping, random DNA mutations are more likely to create some amino acid substitutions than others. This implies that some amino acids are more likely to be tested by mutation as substitutes than others (Fig. 2). However, the situation for amino acid mutations is a bit more flexible than for codon mutations because multiple codons specify most amino acids. Careful inspection of the listing of codons with their associated amino acids shows that each amino acid can convert to 6–13 out of the 19 other amino acids (>1/3 of the possible interconversions, on average). For example in the upper left corner of Fig. 2, the phenylalanine (F) to methionine (M) interconversion (colored white) is not allowed by a single point mutation but all other interconversion of the five amino acids are allowed (colored gray).

Some amino acids substitutions are more likely to be harmless than others due to similar physicochemical properties (e.g., [31]). The essence of the mutational robustness in the codon-to-residue mapping is that these more harmless mutations are more likely to result from a single point mutation or from a more probable mistranslation. In the same way, mutations that are more likely to be harmful are less likely to occur from a single point mutation, instead requiring two or three point mutations in the same codon, a much rarer occurrence [13, 14, 18]. Which amino acid pairs are more substitutable is an important factor in understanding the genetic code, so it would significantly increase the meaningfulness of the codon table if some way were found to express it.



**Fig. 2** Point mutation matrix. If a codon from the y-axis (codons are read from this axis as shown in Fig. 1b intersects with one from the x-axis at a position on the matrix crossed by a *diagonal blue line*, this codon to codon change is “allowed” by a single nucleotide point mutation, i.e., a change in just one of the three codon letters will take the x-axis codon to the y-axis codon or vice versa. If an amino acid from the y-axis intersects with one from the x-axis at a *gray square*, this amino acid to amino acid change is also allowed by a single nucleotide point mutation. Biologically, this means that a given amino acid cannot change to any arbitrary amino acid in one generation (multiple mutations in the same codon are extremely rare). Instead, changes are confined according to this table. For example the phenylalanine (F) to methionine (M) interconversion (*colored white*) is not allowed by a single point mutation—all other interconversions of the five amino acids in

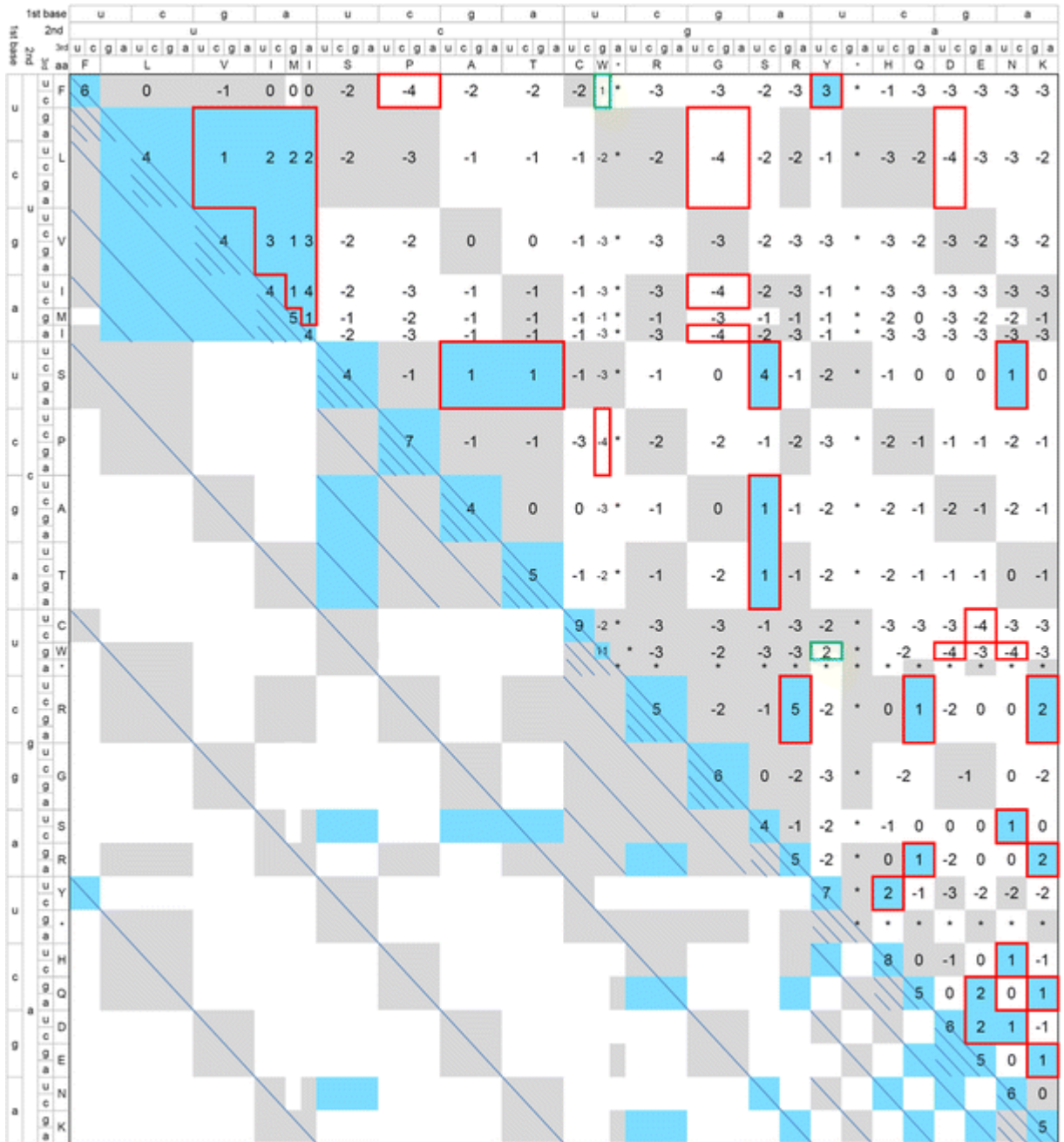
the upper left corner are allowed (*colored gray*). Note that this upper left corner has mostly allowable amino acid to amino acid changes whereas the lower right corner has more of a checkerboard appearance. This is because there are fewer amino acids specified by codons with U in the second base position (upper left) relative to codons with A as the second base (lower right). Also note that the diagonal lines form a fractal pattern with three levels. This is because the ordering of the bases in the three different positions follows a consistent order that is permuted hierarchically to construct the x- and y-axes. In this case, the third base position is permuted most frequently, followed by the first base position and the second base position (though any hierarchy of permutation with a consistent ordering would produce an identical fractal-like pattern)

---

	F	L	V	I	M	S	P	A	T	C	W	R	G	Y	H	Q	D	E	N	K
F	6	0	-1	0	0	-2	-4	-2	-2	-2	1	-3	-3	3	-1	-3	-3	-3	-3	-3
L	0	4	1	2	2	-2	-3	-1	-1	-1	-2	-2	-4	-1	-3	-2	-4	-3	-3	-2
V	-1	1	4	3	1	-2	-2	0	0	-1	-3	-3	-3	-1	-3	-2	-3	-2	-3	-2
I	0	2	3	4	1	-2	-3	-1	-1	-1	-3	-3	-4	-1	-3	-3	-3	-3	-3	-3
M	0	2	1	1	5	-1	-2	-1	-1	-1	-1	-1	-3	-1	-2	0	-3	-2	-2	-1
S	-2	-2	-2	-2	-1	4	-1	1	1	-1	-3	-1	0	-2	-1	0	0	0	1	0
P	-4	-3	-2	-3	-2	-1	7	-1	-1	-3	-4	-2	-2	-3	-2	-1	-1	-1	-2	-1
A	-2	-1	0	-1	-1	1	-1	4	0	0	-3	-1	0	-2	-2	-1	-2	-1	-2	-1
T	-2	-1	0	-1	-1	1	-1	0	5	-1	-2	-1	-2	-2	-2	-1	-1	-1	0	-1
C	-2	-1	-1	-1	-1	-1	-3	0	-1	9	-2	-3	-3	-2	-3	-3	-3	-4	-3	-3
W	1	-2	-3	-3	-1	-3	-4	-3	-2	-2	11	-3	-2	2	-2	-2	-4	-3	-4	-3
R	-3	-2	-3	-3	-1	-1	-2	-1	-1	-3	-3	5	-2	-2	0	1	-2	0	0	2
G	-3	-4	-3	-4	-3	0	-2	0	-2	-3	-2	-2	6	-3	-2	-2	-1	-2	0	-2
Y	3	-1	-1	-1	-1	-2	-3	-2	-2	-2	2	-2	-3	7	2	-1	-3	-2	-2	-2
H	-1	-3	-3	-3	-2	-1	-2	-2	-2	-3	-2	0	-2	2	8	0	-1	0	1	-1
Q	-3	-2	-2	-3	0	0	-1	-1	-1	-3	-2	1	-2	-1	0	5	0	2	0	1
D	-3	-4	-3	-3	-3	0	-1	-2	-1	-3	-4	-2	-1	-3	-1	0	6	2	1	-1
E	-3	-3	-2	-3	-2	0	-1	-1	-1	-4	-3	0	-2	-2	0	2	2	5	0	1
N	-3	-3	-3	-3	-2	1	-2	-2	0	-3	-4	0	0	-2	1	0	1	0	6	0
K	-3	-2	-2	-3	-1	0	-1	-1	-1	-3	-3	2	-2	-2	-1	1	-1	1	0	5

**Fig. 3** The BLOSUM62 matrix [11, 22]. Numbers indicate how likely homologous positions in different proteins are occupied by the two (or one, for diagonal terms) amino acids specified by the matrix indices. For example, it is more likely to find F (phenylalanine) and L (leucine) at homologous positions than F and V (valine), since the F–L value is 0 while the F–V value is –1 (values are determined using databases of homologous proteins and are rounded off to the nearest integer). Interestingly, not all diagonal terms have the same value since some amino acids, e.g., W (tryptophan), are much more highly conserved thus more likely to be found at homologous positions. Here we have colored positive values *blue* and colored the lowest value (–4) *red*

---



**Fig. 4** Superimposed BLOSUM 62 and point mutation matrices. The mutational robustness pattern is evident from the prevalence of values boxed in red (39) relative to those boxed in green (2), which are exceptions to this pattern. Red boxes surround blue shaded regions that are off-diagonal positive (favorable) BLOSUM62 substitution values, all but two of which are also within allowed portions of the mutation matrix. Red boxes also surround the most negative BLOSUM 62 values (i.e. -4), all nine of which are also within not-allowed portions of the mutation matrix

Amino acid substitution matrices such as the well-known PAM250 and BLOSUM62 (Fig. 3) matrices have been developed to express mutability and are heavily used in sequence alignment tasks. A synthesis that integrates substitution matrices and the codon table would be more meaningful than either alone. To help reveal the genetic code's mutational robustness we can superimpose two matrices: the BLOSUM62 matrix (Fig. 3), and a point mutation matrix (Fig. 2). The resulting superposition is shown in Fig. 4.

As can be seen, the positive numbers (showing substitutability, see Fig. 3) are clustered in regions associated with single nucleotide mutations with only two exceptions, indicated by green boxes in Fig. 4. These exceptions both involve the relatively uncommon and biosynthetically costly amino acid tryptophan, which is specified by only a single codon. Depending on which substitutability matrix is used for superposition, there will be different exceptions to the way in which the superimposed matrices reveal mutational robustness. One common argument against code optimization is that matrices used to determine mutational robustness are themselves contaminated with the genetic code, such that amino acids which can exchange by a single point mutation frequently substitute for one another and thus “contaminate” the substitutability matrices. To avoid this problem, substitutability matrices that are not based on observed substitution frequencies are possible instead, as we show for two such matrices (see Supplemental Materials, Fig. S1 [45] and Fig. S2 [51]).

In addition to the positive numbers (relatively favorable substitutions) clustering in allowed regions of the point mutation matrix, the least favorable substitutions (with a value of  $-4$  in the BLOSUM62 matrix, see Fig. 4) are exclusively found in the non-allowed regions. Thus, the two aspects of genetic code mutational robustness—maximizing allowed point mutations for favorable amino acid substitutions and minimizing allowed mutations for unfavorable substitutions—are both clearly seen in the superposition of Fig. 4.

Other aspects of genetic code optimization, such as its ability to embed additional information into coding sequences [24], may be more difficult to visualize. Furthermore, it is clear that there are multiple overlapping codes, including a code for specifying amino acids, a code for specifying alternative amino acids, a code for mRNA folding and processing, and a code for controlling speed of translation (hence protein folding). For multiple overlapping codes to exist within the same DNA sequence would seem to require that each code contain inherent redundancy, otherwise it would be difficult or impossible to keep one code sufficiently stable while evolving other codes expressed by the same sequence. Thus, the codon table itself should gracefully express such redundancy.

The overlapping nature of these codes means that the various levels of encoding must be mutually compatible. While the interdependent nature of these codes may be difficult to represent, a number of alternatives and other visualizations to the standard codon table have been proposed [26, 28, 30, codon wheels like

<http://en.wikipedia.org/wiki/File:GeneticCode21-version-2.svg>, 3, 9, 16, 25]. The superposition in Fig. 4—and shown as a 3-D rendering in Fig. S3—hints at one new method to depict a broader concept of the genetic code.

---

## Conclusion

The motivation for constructing a broader, more accurate understanding of the genetic code has two pragmatic aspects. First, accurate mental models [17] help protect their holders from false beliefs. The more accurate a mental model, the better it can support clarity in a field when challenged by pseudoscience and erroneous arguments. And second, strategic direction of a field can be adversely affected both directly and indirectly by unnecessarily simplistic and inaccurate models. Direct effects arise from the biases of practitioners within the field. Indirect effects are imposed by availability of governmental, commercial, and other resources.

The simple relationship between RNA codons and amino acids will never be adequate to explain the complex molecular relationships that embody the genetic code. An ideal representation of the genetic code continues to elude us today. The standard codon table, although a brilliant advance, now helps obscure this fact. A unifying map that does justice to the code, including but not limited to the incomplete picture provided by the standard codon table, would be a significant benefit. In this article, we have examined the limitations of the codon table as a representation of the genetic code. In a subsequent article we hope to explore an improved representation with properties not present in the ordinary codon table.

It has been said, “No mere tool devised by humans has the complexity of representation found in the genome” [2]. If the codon table is to truly fulfill its proper role as the primary expression of the genetic code, it must be improved to communicate as much as possible about the genetic coding principles.

**Acknowledgments** We thank Thomas Wilhelm (Institute of Food Research, Norwich, UK) and Svetlana Nikolajewa (Hans-Knoell-Institute, Jena, Germany) for a critical reading of this article. This does not imply full agreement with the contents.

---

## Supplementary Material

### S1. Supplemental Figures



**Fig. S1** Superimposed EX75 matrix (Stoltzfus and Yampolsky 2007) and point mutation matrix. The mutational robustness of the genetic code is evident in the prevalence of red boxes relative to green boxes. The 31 most positive ( $\geq 7$ ) and most negative ( $\leq -11$ ) numbers have been outlined with boxes, red for the 24 cases that conform with the mutational robustness pattern, and green for the 7 cases that are exceptions, all but 1 of which are found in the alanine and serine columns and rows. Interestingly, the EX matrices are constructed in large part from alanine scanning experiments for which alanine is the “destination” amino acid, and its alanine data shows a negative correlation with the substitutability of alanine measured by many other matrices (e.g. (Woese, Dugre et al. 1966)). Serine is the second most common destination amino acid in the EX matrix construction. Blue shaded squares are in the same location as in Fig. 4 to facilitate comparison.



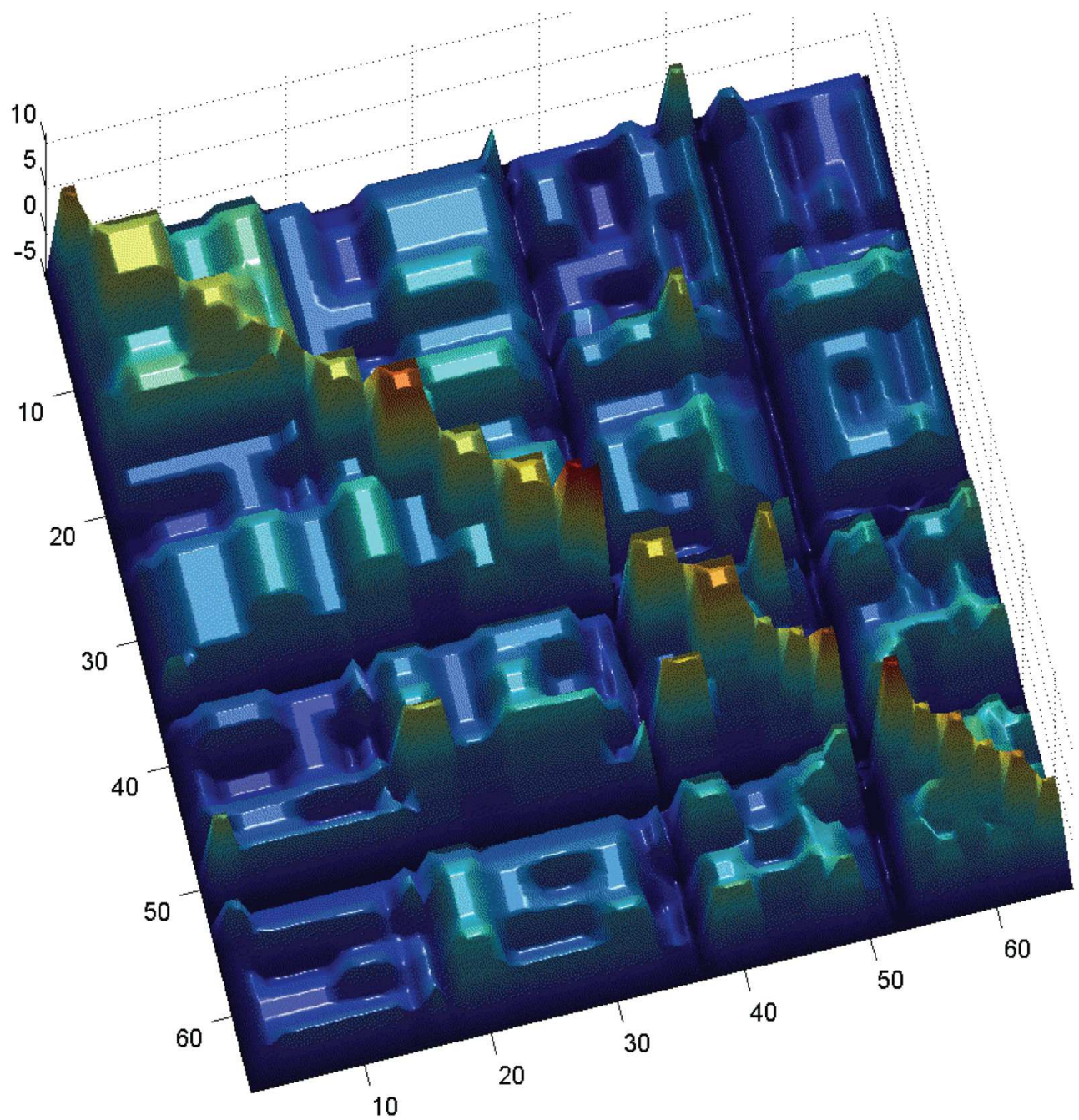
Berleant et al Revised MS-EM542 Fig. S2

1st base		u				c				g				a													
1st base	2nd	u	c	g	a	u	c	g	a	u	c	g	a	u	c	g	a										
	3rd	u	c	g	a	u	c	g	a	u	c	g	a	u	c	g	a										
	aa	F	L	V	I	M	I	S	P	A	T	C	W	*	R	G	S	R	Y	*	H	Q	D	E	N	K	
u	u	F	6	1	-1	0	0	0	-4	-4	-4	-3	-2	2	*	-5	-6	-4	-5	-1	*	-1	-4	-6	-5	-4	-5
c	u	L	3	1	2	2	2	-3	-2	-2	-2	-3	-2	*	-3	-6	-3	-3	-2	*	-3	-3	-5	-4	-4	-3	
g	u	V	2	2	2	1	2	-2	-1	-1	-1	-1	-1	4	*	-3	-4	-2	-3	-2	*	-3	-3	-4	-3	-3	
a	u	I	3	2	3	-3	-2	-2	-1	-1	-2	-3	*	-4	-5	-3	-4	-2	*	-3	-3	-5	-4	-4	-3		
u	c	M	4	1	3	-2	-2	-2	-1	-2	-2	*	-3	-5	-2	-3	-1	*	-3	-2	-4	-3	-3	-3	-3		
c	u	I	3	-3	-2	-2	-1	-2	-3	-4	-2	*	-3	-3	-4	-2	*	-3	-3	-5	-4	-4	-3	-3			
g	u	S	1	0	0	0	-1	-4	-1	-1	1	-1	-3	*	-1	-1	-1	-1	-1	-1	-1	-1	0	-1			
a	u	P	7	-1	-1	-4	-5	-2	-2	0	-2	-4	*	-2	-2	-2	-1	-2	-1	-2	-1	-2	-1				
u	c	A	1	0	-1	-5	-2	0	0	-2	-4	*	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1					
c	u	T	2	-2	-5	-1	-2	0	-1	-3	*	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1						
g	u	C	11	-2	-3	-3	-1	-3	-1	*	-3	-4	-4	-4	-3	-4											
a	u	W	14	*	-3	-5	-4	-3	3	*	-2	-7	-6	-5	-5												
u	c	R	4	-2	-1	4	-3	*	0	1	-2	-1	-1	2													
c	u	G	6	-1	-2	-5	*	-3	-1	-1	0	-1															
g	u	S	1	-1	-3	*	-1	-1	-1	-1	0	-1															
a	u	R	4	-3	*	0	1	-2	-1	-1	2																
u	c	Y	7	*	2	-3	-4	-4	-2	-3																	
c	u	H	5	0	-1	-1	0	0																			
g	u	Q	2	0	1	0	1																				
a	u	D	4	2	1	-1																					
u	c	E	3	0	0																						
c	u	N	3	0																							
g	u	K	2																								

**Fig. S2** Superimposed PAM74-100 matrix and point mutation matrix (Benner, Cohen et al. 1994). Again, the mutational robustness of the genetic code is evident in the prevalence of red boxes relative to green boxes. Here, values  $\geq 0$  and  $\leq -5$  are boxed, red for the 60 cases that conform with the mutational robustness pattern, and green for the 9 cases that are exceptions (if only values  $\geq 1$  and  $\leq -5$  were boxed the ratio of red to green boxes would be 45 to 3, i.e. a ~2-fold better red-green ratio, but 75% less red boxes). Blue shaded squares are in the same location as in Fig. 4 to facilitate comparison.

Berleant et al Revised MS-EM542 Fig. S3

(see next page)



**Fig. S3** A 3-D version of Fig. 4, in which height represents the BLOSUM62 values. As in Fig. 4, the upper left corner is the U-second-position corner, which has all positive numbers (higher heights) with the exception of mutations to or from phenylalanine. The two pairs of deep troughs represent mutations to or from the stop codons, which have been arbitrarily set to a BLOSUM62 value of -5.

## References for section S1

- Benner, S. A., M. A. Cohen, et al. (1994). "Amino acid substitution during functionally constrained divergent evolution of protein sequences." *Protein Eng* **7**(11): 1323-32.
- Stoltzfus, A. and L. Y. Yampolsky (2007). "Amino acid exchangeability and the adaptive code hypothesis." *J Mol Evol* **65**(4): 456-62.
- Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M., & Saxinger, W. C. (1966). On the fundamental nature and evolution of the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology*, **31**, 723–736.

## S2. Analysis of common conceptions of the “Genetic Code”

The first three pages of results from a recent Google web search engine query (search string: “genetic code”) revealed the following characterizations of the genetic code concept. These were placed below into five categories labeled with roman numerals.

### *I. The genetic code as envisioned by the codon table. This is the most prevalent characterization.*

1. The set of codons and the amino acids they make ([www.abc.net.au/science/slab/genome2001/glossary.htm](http://www.abc.net.au/science/slab/genome2001/glossary.htm)).
2. The base triplets that specify the 20 different amino acids ([www.kumc.edu/gec/gloss.html](http://www.kumc.edu/gec/gloss.html)).
3. The mapping between the set of 64 possible three-base codons and the amino acids or stop ([www.bsccs.org/onco/glossary.htm](http://www.bsccs.org/onco/glossary.htm)).
4. The code by which a nucleotide sequence is translated into an amino acid sequence. Each three nucleotide triplet constitutes a codon; the 64 codons correspond to 20 amino acids and to signals for the initiation and termination of transcription ([www.genpromag.com/Glossary~LETTER~G.html](http://www.genpromag.com/Glossary~LETTER~G.html)).
5. Each amino acid building block of a protein is specified by the order of nucleotides (A,C,T and G) in the gene for that protein. Three adjacent nucleotides, called a codon, are required to specify one amino acid. The genetic code can be displayed in a table that translates each of the 64 possible triplet codons into an amino acid. There are 64 possible combinations resulting from having one of four nucleotides in each of three possible positions in the codon ( $4 \times 4 \times 4 = 64$ ) ([www.cgm.northwestern.edu/glossary.htm](http://www.cgm.northwestern.edu/glossary.htm)).
6. The set of correspondences between nucleotide pair triplets in DNA and amino acids in protein ([depts.washington.edu/~genetics/courses/genet372/w2000Terms.html](http://depts.washington.edu/~genetics/courses/genet372/w2000Terms.html)).
7. The base-pair information that specifies the amino acid sequence of a polypeptide ([www.modernhumanorigins.com/g.html](http://www.modernhumanorigins.com/g.html)).
8. Translation specification, which establishes the relationship between the nucleotide sequence in a gene and the amino acid sequence in a protein ([www.the-mwg.com/html/glossary/glossary\\_overview.shtml](http://www.the-mwg.com/html/glossary/glossary_overview.shtml)).
9. The “language” of the genes, dictating the correspondence between nucleotide sequence in DNA and amino acid sequence in proteins; a series of 64 different three-

- nucleotide sequences or triplets (each such triplet is called a codon); except for three “stop” signals, each codon corresponds to one of the 20 amino acids ([www-hsc.usc.edu/~dconti/notes/genetic\\_terms.htm](http://www-hsc.usc.edu/~dconti/notes/genetic_terms.htm)).
10. The set of codons in DNA or mRNA. Each codon is made up of three nucleotides which call for a unique amino acid. For example, the set AUG (adenine, uracil, guanine) calls for the amino acid methionine. The sequence of codons along an mRNA molecule specifies the sequence of amino acids in a particular protein ([www.food.gov.uk/science/ouradvisors/toxicity/cotmeets/49737/49750/49831](http://www.food.gov.uk/science/ouradvisors/toxicity/cotmeets/49737/49750/49831)).
  11. Used to translate the message coded in the gene into a protein. One sequence of 3 nucleotides (codon) corresponds to one amino acid (of the protein) ([www.genethon.fr/php/layout.php](http://www.genethon.fr/php/layout.php)).
  12. The language in which DNA's instructions are written. It consists of triplets of nucleotides, with each triplet corresponding to one amino acid in a protein or to a signal to start or stop protein production ([www.nigms.nih.gov/news/science\\_ed/genetics/glossary.html](http://www.nigms.nih.gov/news/science_ed/genetics/glossary.html)).
  13. The sequence of nucleotides, coded in triplets (codons) along the mRNA, that determines the sequence of amino acids in protein synthesis. The DNA sequence of a gene can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the amino acid sequence ([www.hgsc.bcm.tmc.edu/docs/HGSC\\_glossary.html](http://www.hgsc.bcm.tmc.edu/docs/HGSC_glossary.html)).
  14. Description: Information contained in DNA molecules, recorded by the sequence of the four bases (A, T, G and C) that form the "letters" of this alphabet. Source: Specialized encyclopedia and dictionaries Description: Sequence of nucleotides, coded in triplets (codons) along the mRNA, determining the sequence of amino acids in the production of protein. The four letters of the DNA alphabet (A, C, G, T) form 64 triplets or codons ([europa.eu.int/comm/research/biosociety/library/glossarylist\\_en.cfm](http://europa.eu.int/comm/research/biosociety/library/glossarylist_en.cfm)).
  15. The sequence of nucleotides (building blocks) in the DNA molecule of a chromosome that specifies the amino acid sequence in the synthesis of proteins. It is the basis of heredity ([www.internal.schools.net.au/edu/lesson\\_ideas/dinosaurs/glossary.html](http://www.internal.schools.net.au/edu/lesson_ideas/dinosaurs/glossary.html)).
  16. The way in which the information carried by the DNA molecules determines the arrangement of amino acids in the proteins synthesized by the cells. Each of the 20 amino acids found in proteins is represented by 1 or more units of 3 consecutive nucleotide bases (ie, codons) in the mRNA and in the DNA from which the mRNA is derived. All living organisms and viruses use the same genetic code ([www.niaaa.nih.gov/publications/arh26-3/165-171.htm](http://www.niaaa.nih.gov/publications/arh26-3/165-171.htm)).
  17. The genetic code is a set of rules, which maps DNA sequences to proteins in the living cell, and is employed in the process of protein synthesis. Nearly all living things use the same genetic code, called the standard genetic code, although a few organisms use minor variations of the standard code ([en.wikipedia.org/wiki/Genetic\\_code](http://en.wikipedia.org/wiki/Genetic_code)).
  18. The mechanism by which genetic information is stored in living organisms. The code uses sets of three nucleotide bases (codons) to make the amino acids that, in turn, constitute proteins ([www.kurlama.com/glossary/g.html](http://www.kurlama.com/glossary/g.html)).
  19. The instructions in a gene that tell the cell how to make a specific protein. A, T, G, and C are the "letters" of the DNA code; they stand for the chemicals adenine,

thymine, guanine, and cytosine, respectively, that make up the nucleotide bases of DNA. Each gene's code combines the four chemicals in various ways to spell out 3-letter "words" that specify which amino acid is needed at every step in making a protein ([genencordev.zoomedia.com/wt/gcor/glossary](http://genencordev.zoomedia.com/wt/gcor/glossary)).

## ***II. The genetic code as instructions for protein synthesis***

1. The nucleotide sequence of a DNA molecule (or, in certain viruses, of an RNA molecule) in which information for the synthesis of proteins is contained ([ppathw3.cals.cornell.edu/glossary/Defs\\_G.htm](http://ppathw3.cals.cornell.edu/glossary/Defs_G.htm)).
2. Information carried by the DNA molecules that decides the physical traits of an offspring. The code fixes the pattern of amino acids that build body tissue proteins within a cell ([www.mpssociety.org/lib-glossary.html](http://www.mpssociety.org/lib-glossary.html)).
3. The set of instructions that determines the growth, type, shape, and other characteristics of a living or artificial organism ([www.dakotacom.net/~srooke/glossary.html](http://www.dakotacom.net/~srooke/glossary.html)).

## ***III. The genetic code as the mechanism for storing genetic information***

1. The way in which genetic information is stored in living organisms ([www.perlegen.com/science/dictionary.html](http://www.perlegen.com/science/dictionary.html)).
2. The ordering of nucleotides in DNA molecules that carries the genetic information in living cells ([wordnet.princeton.edu/perl/webwn](http://wordnet.princeton.edu/perl/webwn)).

## ***IV. The genetic code as nucleotide sequence***

1. The DNA sequence of a gene. The genetic code determines the sequence of amino acids in a protein or enzyme, and thus the functions of a living organism ([www.biotech.ca/EN/glossary.html](http://www.biotech.ca/EN/glossary.html)).
2. Exact order (or sequence) of DNA, which makes up genes ([research.marshfieldclinic.org/pmrc/pmrc\\_glossary.asp](http://research.marshfieldclinic.org/pmrc/pmrc_glossary.asp)).

## ***V. Other characterizations of the genetic code***

1. This is carried on chromosomes, which are made up of DNA. Humans have 46 chromosomes. Each chromosome contains many genes which encode various traits ([www-admin.med.uiuc.edu/hematology/Glossary.htm](http://www-admin.med.uiuc.edu/hematology/Glossary.htm)).

---

## **References**

1. Angellotti, M. C., Bhuiyan, S. B., Chen, G., & Wan, X.-F. (2007). CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Research*, 35(Web Server issue), W132–W136.
2. Baltimore, D. DNA is a reality beyond metaphor. Accessed May 20, 2009 from <http://pr.caltech.edu/events/dna/dnabalt2.html>.
3. Biro, J. C. B., Benyó, C., Sanson, A., Szlávecz, G., Fördös, G., Micsik, T., et al. (2003). A common periodic table of codons and amino acids. *Biochemical and Biophysical Research Communications*, 306(2), 408–415.

4. Carter, C. W., Jr. (2008). Whence the genetic code? Thawing the 'frozen accident'. *Heredity*, 100(4), 339–340.
5. Church, G. (2009). Safeguarding biology. *Seed*, 20, 84–86. Accessed March 24, 2009 from [http://seedmagazine.com/content/article/safeguarding\\_biology/](http://seedmagazine.com/content/article/safeguarding_biology/).
6. Cornilescu, G., Delaglio, F., & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR*, 13(3), 289–302.
7. Cortazzo, P., Cervenňansky, C., Marín, M., Reiss, C., Ehrlich, R., & Deana, A. (2002). Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, 293(1), 537–541.
8. Crick, F. H. (1968). The origin of the genetic code. *Journal of Molecular Biology*, 38(3), 367–379.
9. Cristea, P. D. (2003). Large scale features in DNA genomic signals. *Signal Processing*, 83, 871–888.
10. Daniels, L. A. (1996). Selenium metabolism and bioavailability. *Biological Trace Element Research*, 54(3), 185–199.
11. Eddy, S. R. (2004). Where did the BLOSUM62 alignment score matrix come from? *Nature Biotechnology*, 22(8), 1035–1036.
12. Frazer, I. (2005). Gene expression system based on codon translation efficiency. US patent application 20050196865 (continuation of International Patent Application No. PCT/AU2003/001200 filed September 15, 2003).
13. Freeland, S. J., & Hurst, L. D. (1998). The genetic code is one in a million. *Journal of Molecular Evolution*, 47(3), 238–248.
14. Freeland, S. J., Knight, R. D., Landweber, L. F., & Hurst, L. D. (2000). Early fixation of an optimal genetic code. *Molecular Biology and Evolution*, 17(4), 511–518.
15. Freeland, S. J., Wu, T., & Keulmann, N. (2003). The case for an error minimizing standard genetic code. *Origins of Life and Evolution of the Biosphere*, 33(4–5), 457–477.
16. Fujimoto, M. (1987). Tetrahedral codon stereo-table. US Patent 4702704.
17. Gentner, D. (1983). *Mental models*. Hillsdale, NJ: L. Erlbaum Associates.
18. Goodarzi, H., Katanfroush, A., Torabi, N., & Hamed, S. N. (2007). Solvent accessibility, residue charge and residue volume, the three ingredients of a robust amino acid substitution matrix. *Journal of Theoretical Biology*, 245(4), 715–725.
19. Goodenbour, J. M., & Pan, T. (2006). Diversity of tRNA genes in eukaryotes. *Nucleic Acids Research*, 34(21), 6137–6146.
20. Heider, J., Baron, C., & Böck, A. (1992). Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. *EMBO Journal*, 11(10), 3759–3766.

21. Heitzer, M., Eckert, A., Fuhrmann, M., & Griesbeck, C. (2007). Influence of codon bias on the expression of foreign genes in microalgae. *Advances in Experimental Medicine and Biology*, 616, 46–53.
22. Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22), 10915–10919.
23. Howard, M. T., Aggarwal, G., Anderson, C. B., Khatri, S., Flanigan, K. M., & Atkins, J. F. (2005). Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons. *EMBO Journal*, 24(8), 1596–1607.
24. Itzkovitz, S., & Alon, U. (2007). The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Research*, 17(4), 405–412.
25. Jiménez-Montaño, M. A., de la Mora-Basáñez, C. R., & Pöschel, T. (1994). On the hypercube structure of the genetic code. In: H. A. Lim & C. A. Cantor (Eds.), *Proceedings of 3rd International Conference on Bioinformatics and Genome Research*.
26. Jiménez-Montaño, M. A. (2004). Applications of hyper genetic code to bioinformatics. *Journal of Biological Systems*, 12, 5–20. Accessed March 24, 2009 from Software: <http://www.uv.mx/ajimenez/>, manual: <http://www.uv.mx/ajimenez/Manual/HGCodeManual.htm>.
27. Keeling, P. J., & Doolittle, W. F. (1996). A non-canonical genetic code in an early diverging eukaryotic lineage. *The EMBO Journal*, 15(9), 2285–2290.
28. Koonin, E. V., & A. S. Novozhilov (2009). Origin and evolution of the genetic code: The universal enigma. *IUBMB Life*, 61(2), 99–111. Accessed March 24, 2009 from [http://arxiv.org/PS\\_cache/arxiv/pdf/0807/0807.4749v2.pdf](http://arxiv.org/PS_cache/arxiv/pdf/0807/0807.4749v2.pdf), esp. Figs 2 and 4.
29. Liu, Z., Reches, M., Groisman, I., & Engelberg-Kulka, H. (1998). The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in *Escherichia coli*. *Nucleic Acids Research*, 26(4), 896–902.
30. Marth, J. D. (2008). A unified vision of the building blocks of life. *Nature Cell Biology*, 10(9), 1015–1016.
31. Mathura, V. S., & Kolippakkam, D. (2005). APDbase: Amino acid physico-chemical properties database. *Bioinformatics*, 1(1), 2–4.
32. Matthaei, H., & Nirenberg, M. W. (1961). The dependence of cell-free protein synthesis in *E. coli* upon RNA prepared from ribosomes. *Biochemical and Biophysical Research Communications*, 4, 404–408.
33. Mazurs, E. G. (1974). *Graphic representations of the periodic system during one hundred years*. USA: University of Alabama Press.
34. Nakamura, M., & Sugiura, M. (2007). Translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts. *Plant Journal*, 49(1), 128–134.



35. Nakamura, Y., Gojobori, T., & Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Research*, 28(1), 292.
36. Nirenberg, M., & Leder, P. (1964). RNA codewords and protein synthesis. the effect of trinucleotides upon the binding of sRNA to ribosomes. *Science*, 145, 1399–1407.
37. Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 47, 1588–1602.
38. Phillips, R., Kondev, J., & Theriot, J. (2008). *Physical biology of the cell*. New York: Garland Science.
39. Rechavi, O., & Kloog, Y. (2009). Prion and anti-codon usage: Does infectious PrP alter tRNA abundance to induce misfolding of PrP? *Medical Hypotheses*, 72(2), 193–195.
40. Sandman, K. E., Tardiff, D. F., Neely, L. A., & Noren, C. J. (2003). Revised *Escherichia coli* selenocysteine insertion requirements determined by in vivo screening of combinatorial libraries of SECIS variants. *Nucleic Acids Research*, 31(8), 2234–2241.
41. Santos, M. A., Ueda, T., Watanabe, K., & Tuite, M. F. (1997). The non-standard genetic code of *Candida* spp.: An evolving genetic code or a novel mechanism for adaptation? *Molecular Microbiology*, 26(3), 423–431.
42. Santos, M. A. S., & Tuite, M. F. (2004). *Extant variations in the genetic code. Chapter 12 in The genetic code and the origin of life*. New York, NY: Kluwer Academic/Plenum.
43. Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., et al. (2008). Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12), 4685–4690.
44. Stadtman, T. C. (1991). Biosynthesis and function of selenocysteine-containing enzymes. *Journal of Biological Chemistry*, 266(25), 16257–16260.
45. Strauss, S. (2009). We need a satisfactory metaphor for DNA. *New Scientist*, 2696(Feb 23), 22. Accessed March 24, 2009 from <http://www.newscientist.com/issue/2696>. Summarizes Metaphor contests and contested metaphors: from webs spinning spiders to barcodes on DNA. Chapter in
46. B. Nerlich, R. Elliott & B. Larson, eds., *Communicating Biological Sciences*, Ashgate, UK (ISBN 978-0-7546-7633-1).
47. Turanov, A. A., Lobanov, A. V., Fomenko, D. E., Morrison, H. G., Sogin, M. L., Klobutcher, L. A., et al. (2009). Genetic code supports targeted insertion of two amino acids by one codon. *Science*, 323(5911), 259–261.
48. Walczak, R., Carbon, P., & Krol, A. (1998). An essential non-Watson–Crick base pair motif in 3' UTR to mediate selenoprotein translation. *RNA*, 4(1), 74–84.

49. White, M. (2007). The G-Ball, a new icon for codon symmetry and the genetic code. Accessed March 24, 2009 from <http://arxiv.org/abs/q-bio/0702056>. Also see <http://www.codefun.com> and [http://www.codefun.com/Index\\_Books\\_Rafiki.htm](http://www.codefun.com/Index_Books_Rafiki.htm).
50. Wilhelm, T., & Nikolajewa, S. (2004). A new classification scheme of the genetic code. *Journal of Molecular Evolution*, 59(5), 598–605.
51. Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M., & Saxinger, W. C. (1966). On the fundamental nature and evolution of the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology*, 31, 723–736.