An Ode to the Code: Evidence for Fine-Tuning in the Standard Codon Table

Jed C. Macosko* and Amanda M. Smelser

*Department of Physics, Wake Forest University, Winston-Salem, NC 27109, USA. (*Corresponding author: macoskjc@wfu.edu)*

Abstract

The Standard Codon Table (SCT) records the correlation observed in nature between the complete set of 64 trinucleotide codons and the 20 amino acids plus 3 nonsense (i.e. stop or termination) signals. This table was called a frozen accident by Francis Crick, yet current evidence points to optimization that minimizes harmful effects of mutations and mistranslations while maximizing the encoding of multiple messages into a single sequence. For example, a recent article with the running title "The best of all possible codes?" concluded that "evidence is clear" for the optimized nature of the SCT, and another study found that difficult-to-encode secondary signals are minimized in the SCT. Additionally, the initiating amino acid methionine has been found to minimize the nascent peptide chain's barrier to exit the ribosome. Moreover, the symmetry in the SCT between 4-fold-synonymous and <4-fold synonymous codons has been explained in terms of minimizing mistranslation. In this paper, the hypothesis that the finely tuned optimization of the SCT originates in external intelligence is compared to the hypothesis that its fine tuning is due to the adaptive selection of earlier codes. It is concluded that, in the absence of metaphysical biases against this hypothesis, external intelligence better explains the origin of the SCT. Additionally, this hypothesis prompts lines of inquiry that, 50 years ago, would have accelerated the discovery of the now-known features of the SCT and that, today, can lead to new discoveries.

Key words: genetic code, origin of life, adaptive code, error minimizing code, stereochemical origin, frozen accident, amino acid biosynthesis, coevolution, family non-family symmetry

Introduction

In 1976, Francis Crick and coauthors wrote, "The origin of protein synthesis is a notoriously difficult problem" [1]. Proteins are synthesized based on information contained in mRNA, according to an easily-represented map between RNA trinucleotides and protein building blocks [2]. This map describes the flow of information from mRNA to protein in nearly every organism and is usually called "the genetic code".

Here, the map (Figure 1) is called the Standard Codon Table (SCT) to distinguish it both from the *physical machinery* (Figure 2) that enables this flow of

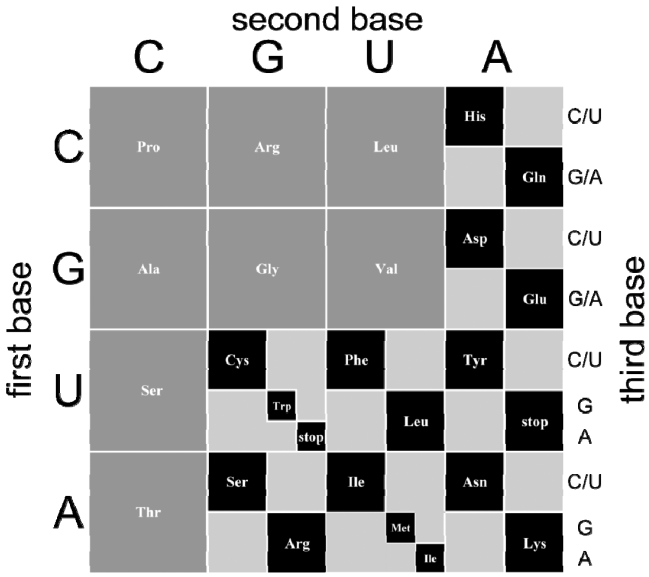

Fig. 1. The Standard Codon Table (SCT) arranged to highlight the family/split-box symmetry. In gray are eight "family" amino acids, specified by four codons each for a total of 32 codons. In black are the other 32 codons: the three stop codons and the codons for the 12 "split-box" amino acids that are coded by three or less codons each. Three amino acids — serine, arginine and leucine — use both family and split-box codons. For purposes of tRNA comparison, the tRNAs that recognize the grey ser, arg, and leu codons are considered family tRNAs and those that recognize the black ser, arg, and leu codons are considered split-box tRNAs.

information and from *additional codes* of secondary signals. These so-called "sub-codes" or "second-layer codes", and the coding machinery itself, are integral parts of the true genetic code, i.e. the full code that starts with the genetic information in DNA and ends with the protein and RNA machines that keep organisms alive [3].

The evolutionary origin of the protein synthesis scheme shown in Figure 2 is what Crick considered a "difficult problem" [1]. There are two parts of this problem: first, how the general coding scheme (Figure 2) originated, and second, how the specific correspondence between trinucleotides and amino acids, i.e the SCT (Figure 1), came about. These two parts are interrelated, but it is helpful at first to consider them separately.

Theories of the Origin of the Standard Codon Table

Currently there are four theories that, alone or in combination, address the origin of the SCT (see review: [4]). First, there is the **frozen accident model**, which takes its name from Crick's suggestion that the SCT was a frozen accident [2]. In other

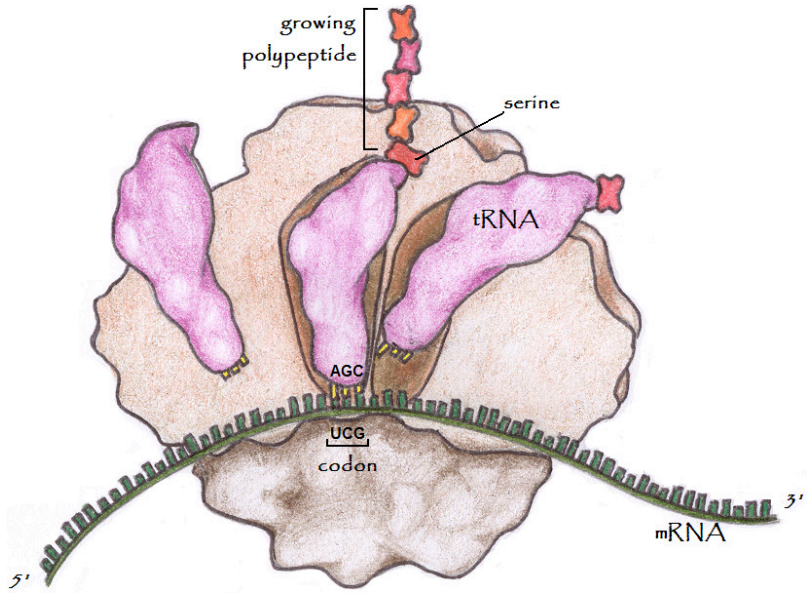

Fig. 2. The tRNAs, shown inside a ribosome, are key pieces of the physical machinery that actualizes the information flow from the mRNA to the polypeptide (protein) chain. This flow follows the SCT; for example, the mRNA letters UCG are recognized by the tRNA that has CGA (as read in the 5' to 3' direction) in its anticodon loop and that carries serine at its opposite end. This example of UCG=serine is shown in Figure 1 (see the grey box labeled "serine" at the intersection of "first base U" and "second base C").

words, neither the mechanism that led to the general coding scheme (Figure 2), nor any other mechanism, dictated the pattern in the SCT (Figure 1). It was purely an accident; the SCT could have ended up with any arbitrary structure. Thus, the current structure does not reveal any information about a past mechanism.

The other three theories all assume that the SCT was not an accident but was formed by a mechanism. By examining the nature of the SCT, one can learn about the mechanism that formed it. The first of these theories is the **error minimization model**. In this model, the SCT was formed by a mechanism that primarily minimized the negative impact of DNA mutations, of mRNA mistranscriptions, and of protein chain mistranslations [5]. Thus, the arrangement of amino acids in Figure 1 is not accidental. For example, once a guanine (G) base in the first codon position and an adenine (A) base in the second position came to represent one of the negatively charged amino acids, then both negatively charged amino acids became encoded with the sequence GAN (where N is any base) so that a mutation in the third position would simply exchange one negatively charged amino acid for another.

Another theory proposes that the origin of the SCT is linked to, or coevolved with, primordial **amino acid biosynthesis** [6]. Several of the 20 amino acids

shown in Figure 1 are synthesized in living cells starting from other amino acids. For example, the negatively charged amino acid, aspartic acid, is known to be a precursor for methionine, threonine, isoleucine, and lysine [7]. These four amino acids are encoded by ANA and ANG codons (see Figure 1), which some take as evidence in favor of this theory [8].

The final theory depends on **stereochemical interactions** between amino acids and their respective trinucleotide codons (Figure 1) or anticodons. This model was popular immediately after the elucidation of the SCT, since it postulated a simple mechanism for the origin of the codon assignments: each codon (or anticodon) had a physical affinity for its respective amino acid, and not for other amino acids [9]. Thus, had this theory proved true, the assignments shown in Figure 1 would have been biochemically predestined by virtue of stereochemical interactions. As it is, the evidence is limited with respect to statistically significant interactions between the codons or anticodons and their respective amino acids. Of the 20 amino acids, only seven (phenylalanine, isoleucine, leucine, histidine, arginine, tyrosine, and tryptophan) show such interactions, and the preference for codon versus anticodon involvement appears random [10].

Of the four theories, error minimization and amino acid biosynthesis are currently favored, though some claim these mechanisms are minor influences compared to the overall frozen accident nature of the SCT [11].

It is important to remember that these four SCT origin theories do *not* explain the origin of the machinery (e.g. Figure 2) that is responsible for converting mRNA information into amino acid sequences. Theories for the origin of the coding machinery are abundant and are generally viewed as extremely speculative (e.g. [12] and reviewer comments). As such, this paper does not address these theories but focuses on just the origin of the codon assignments themselves.

In the next section, we present four studies that describe SCT features that are optimal and are orthogonal, i.e. the optimality of one would not necessarily lead to the optimality of the others. These features are 1) similar amino acids are coded by similar codons thus minimizing the impact of errors, 2) the family/non-family symmetry minimizes mistranslations while maximizing tRNA usage efficiency, 3) the stop codons are related to commonly occurring amino acids in a way that optimizes second-layer codes, and 4) methionine is an optimal initiating amino acid due to its minimized energy for exiting the ribosome.

Orthogonally Optimized Features of the Standard Codon Table

Previous studies [5, 13–16] have compared the optimality of the SCT to those of alternative codon tables in terms of how they mitigate genetic errors by ensuring

that similar amino acids are coded with similar codons (see the "error minimization" theory above). One of these studies in 2000 by Freeland *et al.* determined the most optimized code, given different values of two parameters [15]. The first parameter was the relative likelihood of transitions — A:G or thymine(T):cytosine(C) exchanges — and transversions — A or G exchanging with T or C. The second parameter was the relative impact of mutation as modulated by the power to which the error equation is raised. For most of the intermediate values of these two parameters, the real SCT was the single most optimized codon table — the "best of all possible codes" as this paper's running title suggested. Interestingly, this 100% optimization of the SCT was demonstrated within a restricted set of codon tables. The restricted set reflected the amino acid biosynthesis theory described above. Thus, this paper blended the two favored mechanisms for the origin of the SCT — error minimization and biosynthesis — and quantified a level of optimization that was near or at the global maximum.

Freeland *et al.*'s landmark study tacitly assumes that an optimized code imparts to its owner a selectable advantage over organisms that have not-as-optimized codes. Recent work by Geiler-Samerotte *et al.* helps to answer the question, "What selective effect would a more optimal code have?" [17]. These authors compared the fitness of mutant yeasts expressing a gratuitous protein that misfolded to varying extents. When the protein mostly misfolded and was present at high levels (47,000 copies out of ~40 million total protein molecules per cell, or ~0.1%) the selective disadvantage was 3.2%. Ideally, a selectable disadvantage might be purged from a population when the disadvantage exceeds the inverse of population size, which in yeast is ~10^7 (i.e. 0.00001% when inverted). The authors extrapolate from 47,000 copies to just one misfolded molecule per cell and predict a fitness disadvantage of 0.00008%, that is to say, 8 times greater than the selection threshold. Thus, relative to less optimal codes, any code that results in one less misfolded protein molecule per cell, or even per ~8 cells, can produce a selective advantage. How many less misfolded molecules arise thanks to a "best of all possible" code or a "one in a million" code is still an open question that awaits a direct experimental link between mistranslation rates and misfolding probability.[1]

[1] Interestingly, the Geiler-Samerotte et al. paper nearly provides this experimental link. They state that "random PCR mutagenesis' was performed to generate mutants of the gratuitous protein. 10 mutations out of 238 amino acids were found to cause misfolding. These mutations were: N23I, E32K, G40V, M78V, K101E, I123V, D155G, V163A, Q183H, and S208P. If we assume that these were the complete set of single amino-acid changes that resulted in perceptible misfolding, then the probability that a wrong amino acid causes perceptable misfolding is 10 out of 4522 (i.e. the 238 amino acids multiplied by the 19 possible wrong amino acids at each position). In their study, "perceptible" misfolding appears to be 10%. Thus, for a typical mistranslation rate of 10^{-4} per codon, ~500 codons per protein, and 4×10^7 total proteins per cell, there are >4400 misfolded proteins per

While Freeland *et al.* reported on how the SCT minimizes the impact of errors, another study found an SCT feature that avoids errors in the first place. In 2001, Lim and Curran modeled the specificity of correct codon-anticodon duplex formation during translation [18]. One of the propositions of their model is that, for ribosomes to reject an incorrect duplex, the incorrect duplex must have at least one uncompensated hydrogen bond. This criteria for rejection is problematic when duplexes have a pair of pyrimidines — U (uracil, the RNA equivalent of T) or C — in the wobble position (i.e. third position in codon, first position in anticodon). Pyrimidine bases are smaller than the G and A purine bases and, if they are in the wobble position, they allow certain mismatches in the *second* position to form non-Watson-Crick pairs thereby compensating their missing hydrogen bonds. These mismatches in the second position then fail to be properly rejected and result in a mistranslation event.

This problem of failed rejection nicely explains why 32 codons in the SCT are in "split boxes", and the other 32 are in "family boxes", i.e. the so called family/non-family symmetry of the SCT (see Figure 1). This explanation begins with the observation that the failed rejection problem can be solved by modifying an anticodon's pyrimidine in the wobble position such that it can no longer form a pyrimidine pair. If pyrimidines are modified in this way, then a single anticodon that could have recognized four codons can now only recognize two codons. In other words, there will now need to be one tRNA for the third position pyrimidines, U and C, and another tRNA for the third position purines, A and G.

Lim and Curran's explanation continues with another observation. If each tRNA could recognize four codons apiece, there only would need to be 16 tRNAs for 64 codons. However, these 16 tRNAs could then only encode 16 amino acids. Life requires 20 amino acids and one termination signal, therefore at least some tRNAs must recognize less than four codons (see Figure 3). Conveniently, Lim and Curran showed that there is already a set of tRNAs that must recognize less than four codons — those that are modified to avoid the failed rejection problem.

The choice of which codon boxes in the SCT should be "split" is thus predetermined by the same stereochemistry that determines which mismatches in the second position fall prey to the failed rejection problem. The codons that are susceptible to failed rejection are those with N_1A_2, U_1 or A_1U_2, and U_1 or A_1G_2 — i.e. exactly the split boxes of Figure 1. The symmetry that is observed in the SCT

---

cell. Which means that if a genetic code caused a ~0.02% increase in "wrong" amino acids relative to a different genetic code, it would result in one additional misfolded protein and would therefore, by Geiler-Samerotte *et al.*'s argument, experience negative selection. For comparison, a completely randomized code increases "wrong" amino acids >100 times more, relative to the universal code, than this factor of 0.02%.

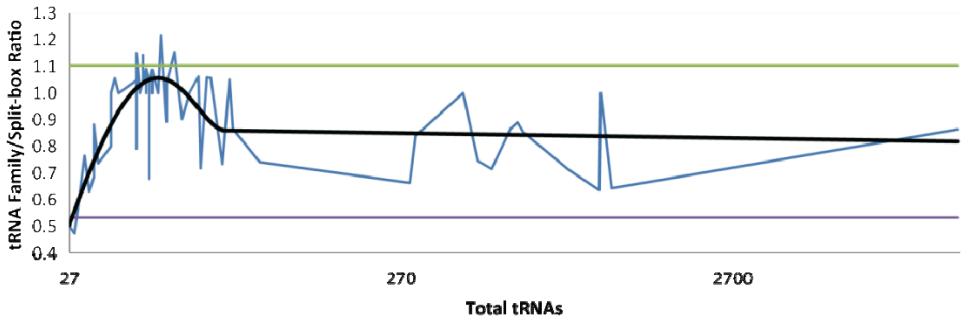

Fig. 3. Family/split-box ratio as a function of total tRNA count (shown in blue, fit with a black line). If each codon had one tRNA, the total tRNA count would be 61 (the three stop codons do not require tRNA) and the family tRNAs to split-box tRNAs ratio would be 32/29 (=1.1, green line). If each amino acid used only one tRNA, the total tRNAs count would be 23 (not 20, since we are double counting arg, leu and ser, as described in the text) and the ratio would be 8/15 (=0.53, purple line). The actual ratio, below 75 total tRNAs, starts at an absolute minimum of 9/18 and climbs to an average that is slightly below 1.1 before settling into an average of 0.85 for organisms with >75 tRNAs (linear fit). The fact that the ratio is below 1.1 for most organisms indicates that tRNA usage is economized via the mechanism described by Lim and Curran (see text).

is not an accident, it is precisely the symmetry one would expect if the SCT was optimized to avoid translation errors, in particular the failed rejection errors due to unmodified pyrimidines in the wobble position.

Itzkovitz and Alon in 2007 described a third remarkable orthogonal advantage of the SCT: the assignments of UAA, UAG, and UGA as stop codons [19]. High frequency codons, such as those coding for aspartic or glutamic acid, can frequently form stop codons if the reading frame shifts. Consequently, translation of a frame-shift error is halted more quickly on average in the real genetic code than in 99.3% of alternative codes, thus saving the cell significant expense. Correlated with this advantage is the SCT's nearly optimal ability to contain secondary signal sequences within the protein-coding sequence, for example, those that encode regulatory and structural protein binding, and splicing sites.

The reason for the correlation between these two advantages is quite simple. Secondary signal sequences are likely to contain all trinucleotide combinations, including UAA, UAG, or UGA, but if any of these three combinations appear as in-frame codons in the protein-coding sequence they will be read as stop codons during translation. However, since, as noted above, UAA, UAG and UGA are frame-shifts of common codons, it is more probable that they can be successfully embedded in the protein-coding sequence. In other words, the first advantage of the SCT (translation of frame shifted sequences stops sooner) leads to the second advantage (secondary signals are embedded more successfully) and vice versa.

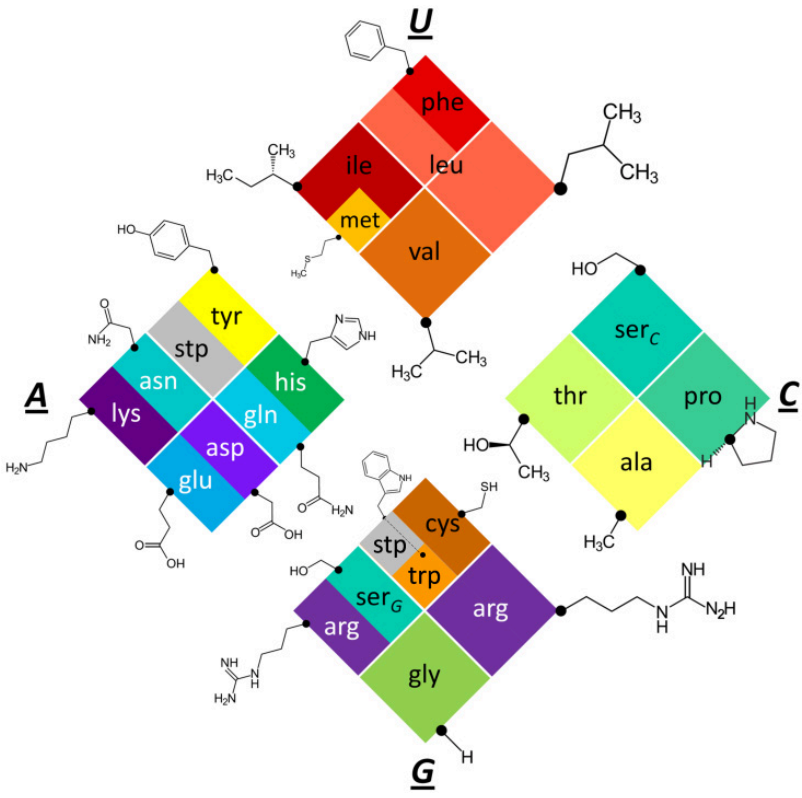

Fig. 4. A new format for displaying the SCT. This version of the new format shows the structure of the 20 amino acid side chains. To identify which trinucleotide codons match which amino acids, follow four steps: 1) Find the quadrant that matches the 2nd base (U=north, G=south, C=east, A=west); 2) Find the square within this quadrant that corresponds to the 1st base (U=north, etc.); 3) Go to the corner of this square that corresponds to the 3rd base (U=north, etc.); 4) Read off amino acid. For example, the AUG codon stands for methionine and has its: 1) second base in the U (north) quadrant 2) first base in the A (west) square 3) third base in the G (south) corner. This new format is useful for showing different patterns in the SCT (see next figure). The rainbow color scheme used here is: most red for most hydrophobic, most blue for most hydrophilic, and grey for the three stop codons. Note, the "family" serine region of the SCT is labeled Ser_C and the "split-box" serine region is labeled Ser_G. Serine is the only amino acid that has codons on the SCT that are not contiguous, i.e. they cannot be connected by single mutations. To go from a Ser_C to a Ser_G codon requires at least two simultaneous mutations.

The fourth orthogonal feature of the SCT is its use of methionine as the initiating amino acid. In 2011 Lim, Curran and Garber devised a novel theory explaining interactions between biomolecules in solution.[2] The lowest barrier to interaction

[2]Lim VI, Curan JF, Garber MB (2012) Hydration shells of molecules in molecular association. A mechanism for biomolecular recognition. J Theo Bio 301:42.

Table 1. Summary of four orthogonally optimized features of the SCT.

Name	Evidence	Extent of optimization
Error impact minimization	Similar amino acids encoded by similar codons	Best possible codes, with restrictions[1]
Error occurrence minimization	Family/ split box symmetry, computer simulation	Specifies symmetry of code[2]
Secondary signal maximization	Stop codons frame shift to common codons	Stop codons vis-a vis common codons
Exit barrier minimization	Initiating methionine has lowest exit barrier	specifies the initiating amino acids

[1] The three restrictions are that all possible codes must have 1) the synonymous codon groupings of the SCT, 2) the stop codons of the SCT, and 3) must not be allowed to change the SCT's groupings of biosynthetically related amino acids.

[2] Placing 32 codons into four-fold synonymous groupings and the other, symmetry-related 32 codons into two-fold synonymous groupings reduces the number of possible codes from 21^{64} (~10^{84}) to $21^{8}\times21^{16}$ (~10^{31}) or a 10^{53}-fold optimization.

results from hydrophobic molecules that present one another with the smallest surface area. A quick inspection of Figure 4 shows that lysine and methionine are the longest, unbranched amino acid residues. Of these two, only methionine is also hydrophobic. Indeed when Lim *et al.* calculated which residue had the lowest interaction barrier, methionine was by far the most optimal.

Besides these four orthogonal features (summarized in Table 1), there are additional SCT features that appear to be orthogonally optimized — three that will be given here as examples. First, the SCT uses fewer codons for rarer and more energetically costly amino acids, thus conserving cellular resources, particularly in mitochondria [20]. Second, it has been shown that frame shifts of the coding and non-coding strands of genes (i.e. protein coding DNA) are more likely to translate into folded proteins than frame shifts of non-genes. In other words, the SCT facilitates the encoding of several proteins in a single region of DNA up to a maximum of six: three on one strand and three on the complementary strand [21]. This high compression of protein data occurs naturally in some viruses that, due the small volume of their capsids, must encode their protein data in their DNA genome as efficiently as possible [22]. Third, the SCT ensures that more common amino acids are less prone to change due to a single base mutation relative to less common ones. This keeps the total number of amino acid changes lower. Interestingly, alternate codon tables that ensure this effect on *both* strands of the DNA are extremely rare, and again the SCT is "one in a million" in this respect [23].

These three additional features are reminders that there are undoubtedly more optimal aspects of the SCT that are awaiting discovery. In the next section, two

theories for the origin of optimality in the SCT will be compared. The first theory depends on the adaptive selection of earlier codes. The second theory depends on the influence of external intelligence. These theories will be evaluated based on whether they plausibly explain the origin of the SCT's optimality in the absence of metaphysical biases. They will also be evaluated based on whether they are conducive to future discoveries of SCT features.

The Origin of Optimality in the Standard Codon Table

The first section of this paper outlined the four theories for the origin of the SCT: frozen accident, error minimization, biosynthesis, and stereochemistry. The second section examined orthogonally optimal features of the code, without specifying models for their origin. In this section, origins are again discussed, but only the origin of the *optimality* of the SCT is considered. Since frozen accident, biosynthesis and stereochemistry are not optimizing mechanisms and produce optimal features only as a collateral effect, they will not be discussed in this section; rather, the error minimization theory will be examined in more detail and compared to the hypothesis that an external intelligence is responsible for the observed optimal features.

Table 1 lists four orthogonally optimal features and the extent of optimization in the SCT due to each one. At first glance, it may seem that one feature — error impact minimization — completely determines any and all optimization in the SCT, since using the error impact criterion alone the SCT was shown to be the most optimal of all possible codes [15]. However, there are three important restrictions placed on the possible codes to which the SCT is compared. First, these other codes must match the SCT in terms of synonymous codons, i.e. the other codes will have the same grey and black boxes shown in Figure 1, but with different amino acids in each box. Second, the other codes must match the SCT in terms of their stop codons, i.e. they all use UAA, UAG, and UGA as stop codons. Third, to construct an alternate code, amino acids cannot swap their positions in Figure 1 with *all* others but only biosynthetically related ones. The four groups of related amino acids used to construct the alternate codes were: 1) Phe, Ser, Tyr, Cys, Trp; 2) Leu, Pro, His, Gln, Arg; 3) Ile, Met, Tyr, Asn, Lys; and 4) Val, Ala, Asp, Glu, Gly.

The SCT is the best of all possible codes within a *specific subset* of possible codes. If one of the three restrictions is relaxed, the SCT is no longer the best of all. For example, the prior work of Freeland *et al.* did not include the third restriction; as a result they found one alternative codon table out of one million attempts that outperformed the SCT in terms of error impact minimization [13]. Interestingly,

the other two restrictions are at least partially set by optimal features discussed earlier (Table 1). Error occurrence minimization [18] partially sets the first restriction—matching synonymous codon boxes — and secondary signal maximization [19] roughly sets the second restriction — UAA, UAG, and UGA stop codons. With two of three restrictions in place, to a first approximation the SCT appears to be at least a "one in a million" code.

The question at this point is: What is the mechanism for the SCT's optimization? It is useful to consider three hypotheses — law, chance, and intelligence [24]. In other words, is the optimization best explained by a predictable, law-like process, by random chance, or by intelligent causation? To distinguish between these choices, it is useful to evaluate them sequentially, beginning with law-like processes. If no law-like processes explain the effect, the probability that chance processes should be considered. Finally, if chance is ruled out based on low probabilities relative to the available time and opportunities, then intelligent causation is by default the best explanation for the effect.

Is there a law that can explain the SCT optimization? Several papers have considered this possibility [4, 11, 25]. For example, if there were primordial organisms that all used different codon tables and if these organisms competed such that only the most fit lineage survived, then by the law-like process of natural selection this lineage would become the last universal common ancestor (LUCA) and its codon table would become the standard for all of life.

Competition between separate lineages with different codes is deemed more likely than a changing code over time within a lineage, where each changed code would need to be backward compatible to the genetic messages of the previous code [2]. Yet despite being more likely, many publications have argued that the laws of competition between lineages cannot explain the SCT's optimization [6, 10, 16, 26–30]. The problem is that if the SCT is "one in a million" there must be a million competing genetic codes in the population of primordial organisms. This problem becomes worse when the optimization of the SCT approaches the "best of all possible codes". In that case, the population of competing codes would need to approach 10^{84} — a ludicrous population size, considering that 10^{84} carbon atoms are a trillion, trillion, trillion times more massive than the earth.

Is chance, then, a reasonable explanation for the SCT's optimization? In 2007 Eugene Koonin invoked the chance hypothesis to explain the complexity of a "translation-replication" system, which would include the SCT, translation components such as shown in Figure 2, and a host of other translation and replication machines [12]. How could a chance occurrence possibly explain even more complexity and optimization than the SCT alone? Koonin's answer is that, if our universe is but one of many in an infinite multiverse, "emergence of highly complex systems by chance is inevitable".

Koonin was criticized by Eric Bapteste in the open access reviewers' comments that accompanied this paper for using a metaphysical argument that "could open a huge door to the tenants of intelligent design". An appeal to an infinite multiverse, which has never been nor can ever be observed, is a poor way to rescue the chance hypothesis from overwhelmingly low probabilities. Better to rule out the chance hypothesis and proceed to the next hypothesis, for even if the particular intelligence responsible for a low probability effect is not known, the general pattern of intelligence producing finely-tuned, optimized effects is well-known and well-studied.

Design is not controversial, but a designer is. All scientists admit that aspects of the universe — and biological systems in particular — conform to various designs that achieve various functions. Yet most scientists reject the possibility that an external intelligence, i.e. a designer, is responsible for the observed design.

There is a persistent, pervasive bias against the design hypothesis, which ensures that even if law and chance fail to explain a biological effect (e.g. the optimization of the SCT), external intelligence will never be considered as an option. However, once this bias is removed, the external intelligence hypothesis becomes the best working hypothesis. Therefore, it should be considered the most viable explanation until a natural mechanism can be found that explains the degree of SCT optimization, or until new data show that the current assessment of optimization is grossly overestimated.

A lingering question is: Why this bias against external intelligence? Possibly, scientists worry that explaining some natural effects via an intelligent force will encourage *all* effects to be explained in this way, thereby dooming the scientific method. This is a reasonable concern. The final section of this paper, therefore, examines the benefits of using external intelligence as a working hypothesis in the specific case of SCT optimization.

Using the Hypothesis of External Intelligence to Guide Discovery

Before the discovery of the SCT in the early 1960's, many researchers assumed that the code would be optimal in some respect. For example, the "diamond" code proposed by George Gamow in 1954 was optimal in its information storage [31]. A chain of N amino acids could be coded by a chain of N+2 mRNA letters, whereas, in the real SCT, N amino acids are specified by 3N mRNA letters. Another pre-SCT code, proposed by Crick, Griffith and Orgel in 1957, was "comma free" and optimal for avoiding frame shifts [32]. Still other codes had interesting mechanisms for automatically correcting errors in translation [33].

With the discovery of the real SCT (see Figures 4 and 5 for a format that is slightly different than Figure 1), two features were immediately recognized: the SCT lacked the host of "nonsense" codons that were required in the comma free codes, and the SCT assigned similar codons to similar amino acids [34–36]. The first feature implied that the physical machinery of the genetic code (e.g. Figure 2) had to be vastly more complex — or more of a random accident — than originally envisioned. The second feature revealed a new type of optimization that was not anticipated, and, surprisingly, was not readily accepted as an optimization. The majority of publications for 30 years seemed intent on explaining *away* this optimization and interpreting the lack of nonsense codons as evidence of randomness rather than complexity (see [37]) of the biosynthetic SCT origin theory via codon expansion, also called "codon capture", where biosynthetically related amino acids capture the codons of amino acids that are already being used in the SCT [38]. In this theory, physiochemical similarities, not biosynthetic pathways, determined how similar codons were assigned to groups of amino acids.

Would SCT research have taken a different tack if external intelligence was considered as its possible source? Would it have taken over 30 years to demonstrate that the obvious pattern of similar amino acids in similar codons confers an impressive level of error impact minimization?

Would other features — secondary signal encoding and error occurrence minimization — have been discovered earlier?

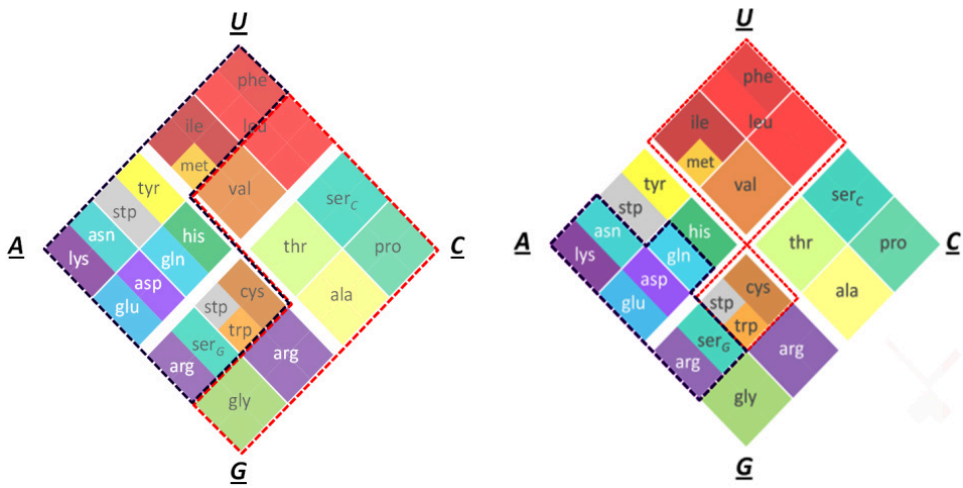

Fig. 5. Example patterns in the standard codon table. Left: Family/split-box symmetry. Right: Hydrophobic (red dashed lines) and hydrophilic (blue dashed lines) amino acids; rankings are the average of five commonly used indices.

At least two papers in the late 1960s suggested that the observed pattern was real optimization and not an artifact of biosynthesis or codon expansion [37, 39]. However, only one of these took an experimental approach and actually tested the SCT against other possible codes, showing that it was more optimal than a random code [37]. This study from 1969 was only cited three times in the 1970s, but gained citations as interest in the optimization of the SCT grew in the late 1980s and into the 1990s. By the time Freeland and Hurst published their "one in a million" paper in 1998, discussion of error impact minimization in the SCT was in full swing.

It is impossible to state unequivocally that optimized features in the SCT would have been discovered and discussed more rapidly in the absence of a bias against external intelligence. However, it is instructive to look at an example from archeology, where external intelligence — i.e. human intelligence — is assumed to account for many features. The Rosetta Stone's discovery in 1799 sparked widespread global interest [40]. Copies were circulated to museums, and each new observation that brought scholars closer to cracking the hieroglyphs was heralded across Europe.

Contrast this scene with the discovery of the SCT. Certainly there was widespread interest, though perhaps shorter lived; an article published three years after the SCT's discovery bore the title "The Genetic Code after the excitement" [41].

The main difference was that the features in the SCT that we now know to be highly optimized were noticed immediately but explained away. Would the discovery today of an intergalactic Rosetta Stone, with the potential to decipher an extraterrestrial language be explained away as an artifact? Certainly not. The bias for or against external intelligence makes all the difference.

There are more features of the SCT that merit examination. Does the proximity in the SCT of biosynthetically related amino acids merely reflect its historical evolution or could this, too, be an optimized feature? Is it significant that the SCT's stop codons would have the weakest codon-anticodon interactions? These and other features will surely be investigated, but the speed at which they will be studied would accelerate if researchers considered the SCT a possible product of external intelligence, with optimized, carefully-engineered features awaiting discovery.

Conclusion

The SCT is by no means the most complex piece of the biological world. On the contrary, its relative simplicity is the reason it has been examined in this paper. Since it is an arrangement of 20 amino acids (and the signal for "stop polymerizing

amino acids") with known properties onto 64 trinucleotides with known properties, it is an ideal test case to examine orthogonal optimized features and to apply the filter of law, chance, and intelligence. If the optimization of the SCT lies between "one in a million" and "the best of all possible codes" as is likely to be the case, the law and chance hypotheses are increasingly untenable and external intelligence becomes the most promising working hypothesis. As new orthogonally optimized features are discovered, the explanatory divide between law and chance on one hand and intelligence on the other becomes more pronounced.

References

1. Crick FHC, Brenner S et al (1976) A speculation on the origin of protein synthesis. Origins of Life and Evolution of Biospheres 7(4): 389–397.
2. Crick FH (1968) The origin of the genetic code. J Mol Biol 38(3): 367–379.
3. Berleant D, White M et al (2009). The genetic code — more than just a table. Cell Biochem Biophys 55(2): 107–116.
4. Knight RD, Freeland SJ et al (1999) Selection, history and chemistry: the three faces of the genetic code. Trends Biochem Sci 24(6): 241–247.
5. Freeland SJ, Wu T et al (2003) The case for an error minimizing standard genetic code. Orig Life Evol Biosph 33(4–5): 457–477.
6. Wong JT (2005) Coevolution theory of the genetic code at age thirty. Bioessays 27(4): 416–425.
7. Lehninger AL, Nelson DL et al (2000) Lehninger principles of biochemistry. Worth Publishers, New York.
8. Wong JT (2007) Question 6: coevolution theory of the genetic code: a proven theory. Orig Life Evol Biosph 37(4–5): 403–408.
9. Woese CR, Dugre DH, et al (1966) On the fundamental nature and evolution of the genetic code. Cold Spring Harb Symp Quant Biol 31: 723–736.
10. Yarus M, Caporaso JG et al (2005) Origins of the genetic code: the escaped triplet theory. Annu Rev Biochem 74: 179–198.
11. Wolf YI, Koonin EV (2007) On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation and subfunctionalization. Biol Direct 2: 14.
12. Koonin EV (2007) The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life. Biol Direct 2:15.
13. Freeland SJ, Hurst LD (1998) The genetic code is one in a million. J Mol Evol 47(3): 238–248.
14. Freeland SJ, Knight RD et al (2000). Measuring adaptation within the genetic code. Trends Biochem Sci 25(2): 44–45.

15. Freeland SJ, Knight RD et al (2000) Early fixation of an optimal genetic code. Mol Biol Evol 17(4): 511–518.
16. Ronneberg TA, Landweber LF et al (2000) Testing a biosynthesis theory of the genetic code: fact or artifact? Proc Natl Acad Sci USA 97(25): 13690–13695.
17. Geiler-Samerotte KA, et al (2011) Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc Natl Acad Sci USA 108(2): 680–5.
18. Lim VI, Curran JF (2001) Analysis of codon: anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. RNA 7(7): 942–957.
19. Itzkovitz S, Alon U (2007) The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res 17(4): 405- 412.
20. Swire J, Judson OP, et al (2005) Mitochondrial genetic codes evolve to match amino acid requirements of proteins. J Mol Evol 60(1): 128–139.
21. Chirico N, Vianelli A et al (2010) Why genes overlap in viruses. Proc Biol Sci 277(1701): 3809–3817.
22. Barrell BG, Air GM et al (1976) Overlapping genes bacteriophage phiX174. Nature 264(5581): 34–41.
23. Bubak M, van Albada G et al (2008). Optimization of asymmetric mutational pressure and selection pressure around the universal genetic code. In: Computational Science-ICCS 2008. Springer, Berlin/ Heidelberg.
24. Dembski WA (1998) The design inference: eliminating chance through small probabilities. Cambridge University Press, Cambridge/New York.
25. Koonin EV, Novozhilov AS (2008) Origin and evolution of the genetic code: The universal enigma. IUBMB Life.
26. Kobayashi K, Furuta H et al (1989) [RNA world, a key step in the origin of life: inspection from the view point of chemical evolution]. Tanpakushitsu Kakusan Koso 34(2): 124–137.
27. Lazcano A, Miller SL (1996) The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. Cell 85(6): 793–798.
28. Soll D, RajBhandary UL (2006) The genetic code-thawing the ‘frozen accident’. J Biosci 31(4): 459–463.
29. Stoltzfus A, Yampolsky LY (2007) Amino acid exchangeability and the adaptive code hypothesis. J Mol Evol 65(4): 456–462.
30. Carter CW Jr (2008) Whence the genetic code? Thawing the ‘frozen accident’. Heredity 100(4): 339–340.
31. Gamow G (1954) Possible relation between deoxyribonucleic acid and protein structures. Nature 173(4398): 318.
32. Crick FHC, Griffith JS et al (1957) CODES /WITHOUT COMMAS. Proc Natl Acad Sci USA 43(5): 416–421.

33. Golomb S, Davey J et al (1963). Synchronization. Comm Sys, IEEE Trans 11(4): 481–491.
34. Matthaei H, Nirenberg MW (1961) The dependence of cell-free protein synthesis in E. coli upon RNA prepared from ribosomes. Biochem Biophys Res Commun 4: 404–408.
35. Nirenberg MW, Matthaei JH (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci USA 47: 1588–1602.
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. Alff-Steinberger C (1969). The genetic code and error transmission. Proc Natl Acad Sci USA 64(2): 584–591.
38. Wu H-L, Bagby S (2005) Evolution of the genetic triplet code via two types of doublet codons. J Mol Evol 61(1): 54–64.
39. Goldberg AL, Wittes RE (1966) Genetic Code: aspects of organization. Science 153(3734): 420–424.
40. Parkinson RB (1999) Cracking the code: the Rosetta Stone and the art of decipherment. British Museum, London.
41. Sadgopal A (1968) The genetic code after the excitement. Adv Genet 14: 325–404.