Quarterly Reviews of Biophysics

cambridge.org/qrb

Energy mapping of the genetic code and genomic domains: implications for code evolution and molecular Darwinism

Horst H. Klump[1], Jens Völker[2] and Kenneth J. Breslauer[2,3]

[1]Department of Molecular and Cell Biology, University of Cape Town, Private Bag, Rondebosch 7800, South Africa; [2]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, 610 Taylor Rd, Piscataway, NJ 08854, USA and [3]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ 08901, USA

Review

Cite this article: Klump HH, Völker J, Breslauer KJ (2020). Energy mapping of the genetic code and genomic domains: implications for code evolution and molecular Darwinism. *Quarterly Reviews of Biophysics* **53**, e11, 1–8. https://doi.org/10.1017/S0033583520000098

Received: 21 August 2020
Revised: 4 October 2020
Accepted: 6 October 2020

Key words:
Codon energy distribution spectrum; codon free energies; codon usage frequency; energy code; interlocking thermodynamic cycles; molecular evolution

Author for correspondence:
Kenneth J. Breslauer,
E-mail: kjbdna@rutgers.edu

Abstract

When the iconic DNA genetic code is expressed in terms of energy differentials, one observes that information embedded in chemical sequences, including some biological outcomes, correlate with distinctive free energy profiles. Specifically, we find correlations between codon usage and codon free energy, suggestive of a thermodynamic selection for codon usage. We also find correlations between what are considered ancient amino acids and high codon free energy values. Such correlations may be reflective of the sequence-based genetic code fundamentally mapping as an energy code. In such a perspective, one can envision the genetic code as composed of interlocking thermodynamic cycles that allow codons to 'evolve' from each other through a series of sequential transitions and transversions, which are influenced by an energy landscape modulated by both thermodynamic and kinetic factors. As such, early evolution of the genetic code may have been driven, in part, by differential energetics, as opposed exclusively by the functionality of any gene product. In such a scenario, evolutionary pressures can, in part, derive from the optimization of biophysical properties (e.g. relative stabilities and relative rates), in addition to the classic perspective of being driven by a phenotypical adaptive advantage (natural selection). Such differential energy mapping of the genetic code, as well as larger genomic domains, may reflect an energetically resolved and evolved genomic landscape, consistent with a type of differential, energy-driven 'molecular Darwinism'. It should not be surprising that evolution of the code was influenced by differential energetics, as thermodynamics is the most general and universal branch of science that operates over all time and length scales.

Table of contents

Introduction **1**
Biophysical origins and evolution of the genetic code 2
Toward a thermodynamic contribution to the genetic code and to molecular Darwinism 2

The genetic code as an energy code **2**
Energy dispersion of the codon/complementary codon, trimeric duplexes 3
Correlations between codon usage frequencies and the stabilities of the duplexes formed by each codon and its antiparallel complementary codon 4
'Evolving' all codons from each other through sequential series of transitions and transversions 4
Correlating trimeric duplex stability with amino acid coding properties 5
Correlations between larger domain DNA energy profiles and higher-order biological functions 6

Concluding remarks **6**
Going forward 7

Introduction

Once the genetic code was deciphered, it was quickly recognized that the code matrix was decidedly nonrandom. Elucidation of the underlying causes of this surprising regularity has been described as 'The Universal Enigma'. Thought leaders such as Francis Crick, Manfred Eigen, Ed Trifonov, and others (see reviews/overviews and references cited therein) (Crick, 1968; Eigen and Winkler-Oswatitsch, 1992; Trifonov, 2000, 2004; Koonin and Novozhilov, 2009), proposed fundamental, physio-chemical frameworks to explain the origin and

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

evolution of the code, including error minimization (Freeland *et al.*, 2003; Novozhilov and Koonin, 2009), stereochemical (Yarus *et al.*, 2005; Polyansky and Zagrovic, 2013; de Ruiter and Zagrovic, 2015), and coevolution theories (Di Giulio, 2004; Wong, 2005). They sought to explain the near singularity of the code across most living organisms, out of 10^{84} alternative possibilities (Koonin and Novozhilov, 2009). To this end, a thermodynamic perspective/framework helps rationalize the origins and evolution of the genomic energy landscape, in which differential energy profiles correlate with differential biological outcomes.

Biophysical origins and evolution of the genetic code

Within the context of an energy-based perspective for the origin and evolution of the genetic code, one can hypothesize that the original ancestral codons were comprised of a family of 'prebiotic' duplexes of sufficient stability to avoid dissociation from their antiparallel, complementary codons. One might reasonably envision that such codon couplets would code for the most ancient amino acids (Miller, 1953; Miller and Urey, 1959; Miller *et al.*, 1976; Trifonov and Bettecken, 1997; Trifonov, 2000, 2004) if given the proper translational machinery. This core of 'prebiotic' 'codon duplexes' could have produced the rest of the code through a sequential series of transition and transversion mutations, the order of which is controlled/regulated/influenced via families of interlocking energy cycles. In this way, one can converge to a nearly singular genetic code out of the potential of more than 10^{84} alternative code tables (Novozhilov *et al.*, 2007; Koonin and Novozhilov, 2009). Such energy-driven convergence of an astronomical number of potential alternative 'states/codes' into a single, unique ensemble 'state/code' is reminiscent of the energy landscape funnel associated with the protein folding phenomenon (Dill and Chan, 1997). Alternatively, one could argue that evolution was 'stuck with' whatever code developed by chance in the primordial environment. From that code, sequential mutations may have worked to drive the code toward evolutionarily traceable changes. Thus, once organisms started using such codes to connect replication (and transcription) to translation, they evolved more rapidly toward a thermodynamically driven code. We prefer the former perspective of an early stage, energy-driven, nonrandom evolutionary shaping of the genetic code through a series of mutations within particularly stable, prebiotic 'codon duplexes', all controlled via families of interlocking energy cycles.

Toward a thermodynamic contribution to the genetic code and to molecular Darwinism

In the classic view of Darwinian evolution (Darwin, 1859), a phenotypical characteristic of a species that imparts a survival advantage in a given environment persists through generations. By contrast, the population of those species variants that lack such an advantageous characteristic disappears over time. This 'survival of the fittest' phenomenon is what gives rise to what Darwin referred to as 'natural selection'.

A necessary correlate to such a view of evolution is that the selective survival of the phenotypically advantaged form of a species yields enrichment within the species of genotypic signatures that correspond to production of (coding for) the advantageous phenotypical characteristics. Within such a framework, one could envision the generational persistence of certain species' characteristics that not only provide selective advantage for survival within a given environment, but which also are associated with particularly stable genetic signatures (codes) that resist alterations and thus enable generational persistence. As such, evolution may result from a mixture of contributions from classic, natural selection, Darwinian theory, as well as from what can be called 'molecular Darwinism' (Eigen, 1976), or 'Watson–Crick Darwinism'. In the latter context, some phenotypical characteristics might persist since they are coded for by more stable domains in the genome, even if such characteristics do not maximize species survival. In the classic evolutionary context, characteristics that provide a survival advantage will generationally persist, with no consideration for the stability of the genotypical signature. The net outcome of species evolution may well reflect contributions from both phenotypical (classic Darwinism) and genotypical (molecular Darwinism) influences.

To gain insight into such a DNA-based perspective, one can map the iconic chemical genetic code in terms of an energy code. Perhaps early evolution of the genetic code can be viewed in terms of the differential stabilities of codon/complementary codon couplets that form antiparallel trimeric duplexes, as opposed to exclusively in terms of the functionality of any gene product.

The genetic code as an energy code

In Table 1, the classic Crick genetic code matrix (Crick, 1968) is elaborated by also listing the stabilizing free energy values for each fully paired codon/complementary codon antiparallel trimeric duplex. These stability parameters were calculated using calorimetrically-derived free energy values previously reported by our labs (Breslauer *et al.*, 1986). The decision to map the genetic code in terms of codon/antiparallel complementary codon energetics is justified on multiple physio-chemical levels. First, trimeric duplexes formed from association of fully complementary, antiparallel, codon pairs, using up to four 'letters', reflect the minimum molecular information units required to code for the diversity of amino acids. Second, as demonstrated by Porschke and Eigen (1971), codon/complementary codon, antiparallel, trimeric duplexes possess the minimum stability required to form a complex of sufficient strength such that the associated species do not spontaneously dissociate into their constituent single-stranded components; a circumstance that makes them susceptible to degradation and thereby loss of coding information. On the other hand, making these interactions still more stable, as in a tetramer or higher chain length code, may well work against optimal rates of translation (see Greive and von Hippel, 2005). In other words, in the current context, increased thermodynamic stability might work against the use of more stable (e.g. longer) codon–complementary codon interactions in 'molecular evolution'. Third, it has been suggested that in primitive primordial molecular machines, translocation events in steps of 3 monomer units (i.e. the size of the codon) correspond to local energy minima that are favored over other translocation step sizes (Aldana *et al.*, 1998; Martinez-Mekler *et al.*, 1999; Aldana-González *et al.*, 2003). It is within this context, that stability data for each antiparallel, codon/complementary codon interaction are shown in Table 1. In this format, one can explore relationships between the differential trimeric duplex stabilities formed from codons

https://doi.org/10.1017/S0033583520000098 Published online by Cambridge University Press

Table 1. Genetic code matrix annotated with trimeric duplex stabilities formed between codons and their corresponding, antiparallel complementary codons

1st	2nd Codon Position: T		C		A		G		3rd
T	3.5	Phe	3.2	Ser	2.8	Tyr	3.2	Cys	T
	3.8	Phe	4.4	Ser	3.2	Tyr	4.6	Cys	C
	3.0	Leu	3.4	Ser	3.0	STOP	3.4	Trp	A
	3.9	Leu	4.5	Ser	3.2	STOP	4.4	Trp	G
C	3.7	Leu	4.3	Pro	3.7	His	4.4	Arg	T
	4.0	Leu	5.5	Pro	4.1	His	5.8	Arg	C
	3.2	Leu	4.4	Pro	3.9	Gln	4.3	Arg	A
	4.1	Leu	5.6	Pro	4.1	Gln	5.6	Arg	G
A	3.3	Ile	3.1	Thr	3.3	Asn	3.1	Ser	T
	3.6	Ile	4.3	Thr	3.7	Asn	4.5	Ser	C
	2.8	Met	3.2	Thr	3.5	Lys	3.2	Arg	A
	3.7	Met	4.4	Thr	3.7	Lys	4.3	Arg	G
G	3.7	Val	4.5	Ala	3.6	Asp	4.3	Gly	T
	4.0	Val	5.7	Ala	4.0	Asp	5.7	Gly	C
	3.2	Val	4.6	Ala	3.8	Glu	4.4	Gly	A
	4.1	Val	5.8	Ala	4.0	Glu	5.5	Gly	G

Genetic code matrix for human mitochondrial DNA annotated with stabilizing free energy values for the trimeric duplexes formed by antiparallel, complementary codons. The free energy values were calculated based on the calorimetrically determined nearest-neighbor dataset reported by Breslauer *et al.* (1986). To reduce the impact of end effects, the normalized, weighted average of the duplex free energies were calculated, with each terminus hypothetically 'sealed' by all possible base pairs/stacks. The net impact of this end effect correction is to dampen the variability between codons, and to reduce the dominance of the central base in the codon trimer caused by it being the only base with two neighbors. Significantly, however, aside from modest compression of the stability range, the rank order of the differential codon duplex stabilities listed, as well as the general correlations noted here remain unaltered, even when no end effect 'correction' is applied. Compilations of the codon free energy data employing any of the other commonly used nearest-neighbor databases results in some numerical and rank order differences**,** reflective of subtle differences in the numerical values assigned to nearest-neighbors in these different databases (e.g. Delcourt and Blake, 1991; Doktycz *et al.*, 1992; SantaLucia *et al.*, 1996; SantaLucia, 1998). That said, the nearest-neighbor data of Ritort and Bustamante, derived from force stretching experiments, exhibit the most concurrence with the relative trends reported here; specifically in terms of the free energy rank ordering of the codons, as well as the codon usage patterns (Huguet *et al.*, 2010). Given Ritort's subsequent assessment of the differential impact of magnesium ion on the nearest-neighbor data, future studies should also consider measurements in a variety of counterion/cation environments (Huguet *et al.*, 2017).

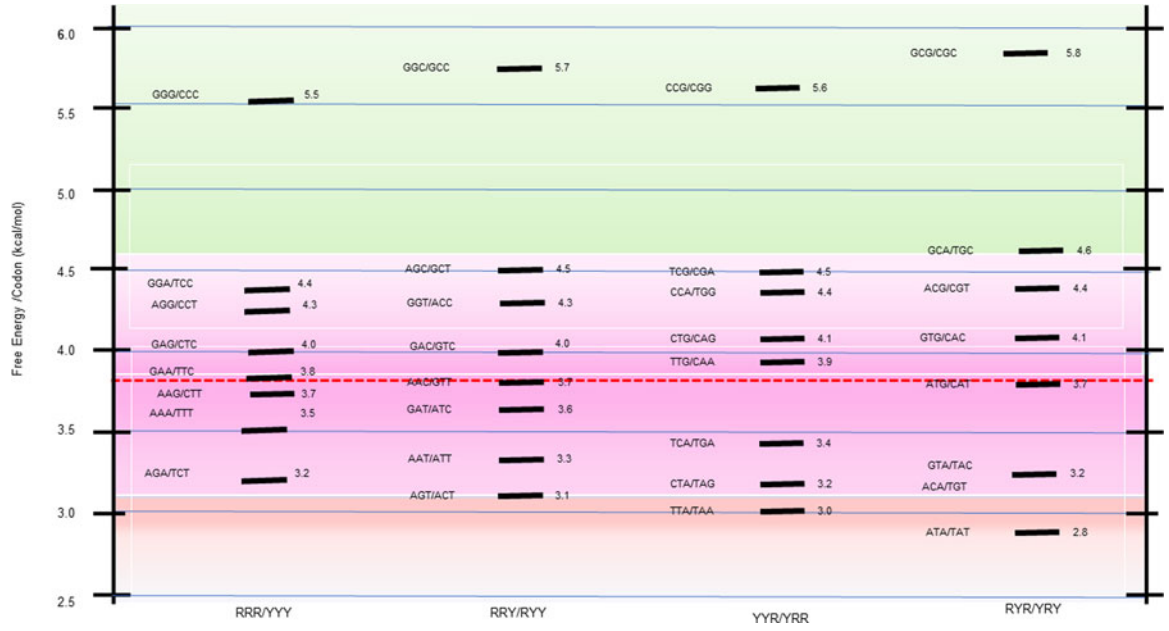
and their corresponding complementary codons and biological observations/outcomes.

The genetic code matrix shown in Table 1 corresponds to the human mitochondria code (Anderson *et al.*, 1981; Breitenberger and RajBhandary, 1985). Note that the smallest interaction free energy (least stable), ΔG, of 2.8 kcal per mole triplet is between the alternating sequences ATA and TAT, and the highest (most stable) ΔG value is 5.8 kcal per mole triplet for the duplex formed by the alternating sequences GCG and CGC.

Energy dispersion of the codon/complementary codon, trimeric duplexes

The 64 possible DNA triplet codons collectively constitute the minimal 'words' of the genetic code. When evaluated as bound to their fully complementary, antiparallel codons, they form 32 trimeric duplexes, each with a calculated stability. As summarized in Scheme 1 heat map, this process reveals a broad stability dispersion of the trimeric duplexes formed by the codon/

https://doi.org/10.1017/S0033583520000098 Published online by Cambridge University Press



Scheme 1. Free energy distribution spectrum for the 32 trimeric duplexes formed by all 64 complementary codons. The stability distribution is color coded as a 'heat map', with the GC-rich most stable family (highest free energy of trimeric duplex formation) highlighted toward the top of the scheme in light green; the next most stable family is highlighted in light purple; and the less stable duplexes relative to the mean are highlighted in light red. The energy spectrum is formatted within four columns that reflect the purine (R)/pyrimidine (Y) sequence patterns designated at the bottom of the scheme.

complementary antiparallel codon. Such significant energy dispersion makes these differential energy profiles information-rich.

Correlations between codon usage frequencies and the stabilities of the duplexes formed by each codon and its antiparallel complementary codon

Comparisons between the codon/complementary codon free energy data and the whole genome codon usage frequencies reported by Futcher and coworkers (Gardin *et al.*, 2014) for yeast *Saccharomyces cerevisiae* reveal an intriguing coupling of properties. By combining the differential stabilities and Futcher's datasets, one observes a near linear correlation between the frequency with which a given codon is used for a particular amino acid and the corresponding codon/complementary codon free energy. To be specific, save for isoleucine, of the 17 out of 20 amino acids for which Futcher reports sufficient data density, we observe that codons with lower free energies (less stable) are used more frequently than codons for the same amino acid with higher free energies (more stable). This coupling of a fundamental physio-chemical property with the outcome of a complex biological process is illustrated for several amino acids in the plots shown in Fig. 1. Such empirical correlations reinforce reductionists' efforts to rationalize complex biology in terms of fundamental chemical principles.

Based on this coupling, one might speculate, that the degeneracy associated with the use of multiple codons of differential stabilities to code for the same amino acid reflects a form of thermodynamic selection; one in which codon energetics is more determinatory of usage frequency than a codon's chemical syntax alone. It will be instructive to probe the extent to which this empirical correlation between codon stabilities and codon usage frequencies is universal across all organisms and genes, as well as to define the biological implications. For now, this correlation provides an example of insights that can be gained by parsing the iconic genetic code in terms of energy differentials. Conversely, one also might posit that the usages of codons reflect biologically relevant features of those DNA sequences containing a statistical overabundance of energetically favorable or unfavorable codons. The altered energy profiles of such DNA sequences relative to a statistically expected distribution of codons/energies may reflect the existence of biological constraints that do not apply to an average sequence. In other words, codon usage that deviates from the average expected distribution (either positive or negative) may reflect altered biological constraints.

'Evolving' all codons from each other through sequential series of transitions and transversions

To illustrate the cycles associated with interconverting codons for the entire genetic code, the 64 codons presented in Table 1 and Scheme 1 can be arranged into a total of eight 'octets'. Each octet is composed of codons located at one of the eight apices of a cube. Each cube corresponds to one of the purine (R)/pyrimidine (Y) sequence patterns designated in Scheme 1. The resultant eight octet cubes are then positioned at the corners of a master scaffolding cube to create a 'hypercube' as shown in Fig. 2. This hypercube illustrates the full cascade of all of codon interconversions via sequential site changes over all codon sequence space.

Note that the eight octet cubes shown in the hypercube of Fig. 2 are inter-cube related by codon transversion mutations, whereas codons within a given octet cube are intra-cube related by transition mutations. These stepwise interconversions create interlocking cycles that allow one to traverse/"evolve" the entire genetic code.

This stepwise generation/"evolution" of all 64 codons via sequential transition and transversion mutations, starting from any codon, may reflect a differential, energy-modulated evolution

https://doi.org/10.1017/S0033583520000098 Published online by Cambridge University Press

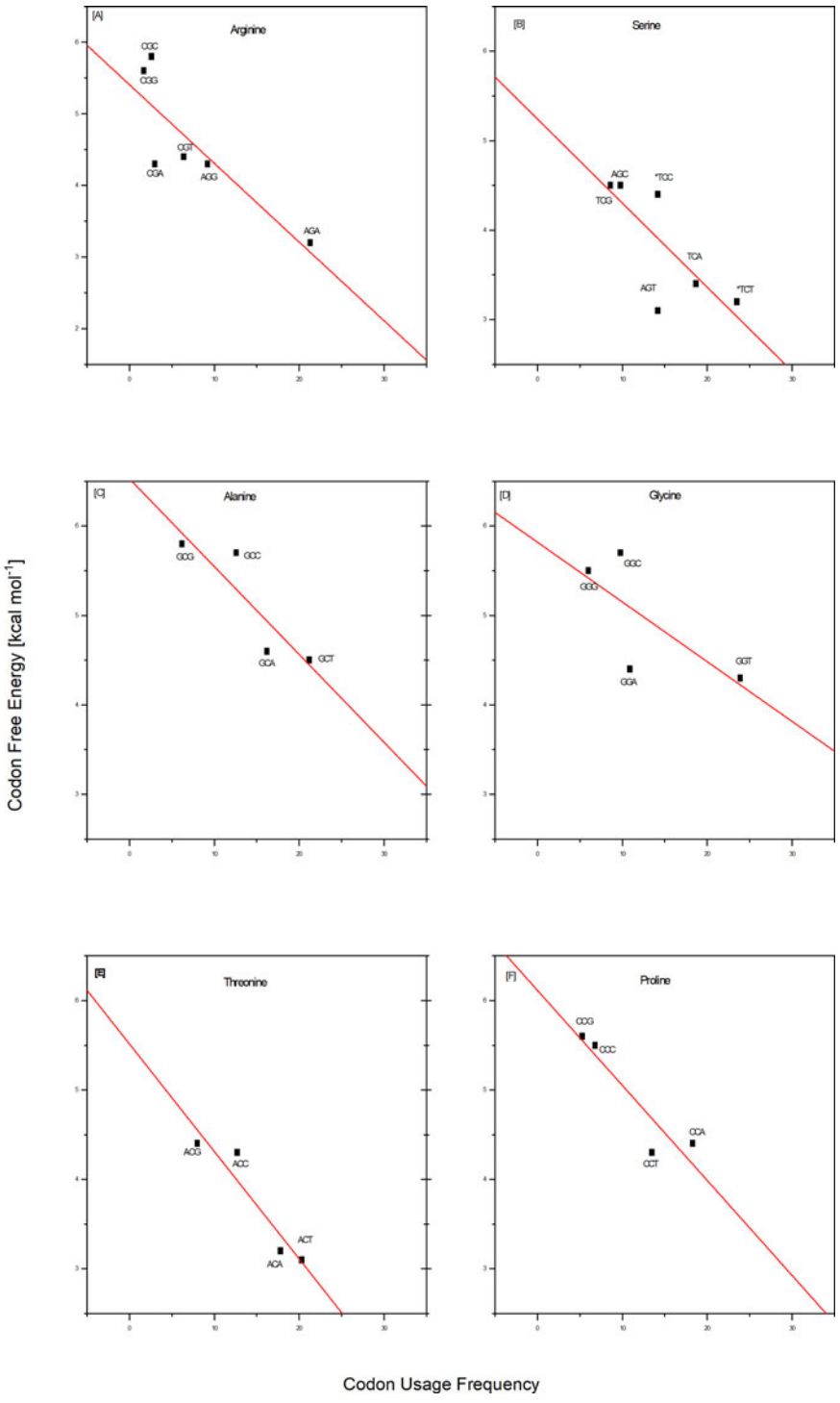


Fig. 1. Empirical correlations between whole genome codon usage frequencies in *S. cerevisiae* taken from the work of Futcher and coworkers (Gardin *et al.*, 2014) and the corresponding codon/complementary codon free energies of this study. Each red line represents a best fit to the equation for a straight line of these two independently derived data sets. The result shown here are for two of the three amino acids encoded by six codons, and for four of the five amino acids encoded by four codons. This selection corresponds to that subset of the amino acids judged most ancient, based on a meta-analysis reported by Trifonov (2000, 2004). With the exception of isoleucine, and the insufficient data density for methionine and tryptophan, all of the amino acids encoded by only two codons also show a preference for higher codon usage frequency that correlates with lower codon free energy. For a thermodynamic argument, one strictly should use a log scale plot for the usage frequency. However, over the small data range assessed here, we have confirmed that one cannot distinguish between linear and log linear, with the log plot simply compressing the data.

of the genetic code. As such, it might correspond to a biophysical basis for what Eigen referred to as 'molecular Darwinism'. At the heart of molecular Darwinism is the generation of genetic variation. Genetic variants can result from DNA sequence alterations at local levels; from rearrangement of DNA segments intragenomically; and by gene transfer of foreign DNA. The hypercube illustrates how local sequence changes can interconvert all 64 codon variants via sequential cascades of transition and transversion steps that Table 1 shows exhibit differential free energies, thereby creating codon variants; a characteristic at the heart of molecular Darwinism.

Correlating trimeric duplex stability with amino acid coding properties

Inspection of the data in Table 1 and Scheme 1 reveals eight exceptionally stable all-GC codons, as defined by the relative stability of the trimeric duplex they each form with their

https://doi.org/10.1017/S0033583520000098 Published online by Cambridge University Press

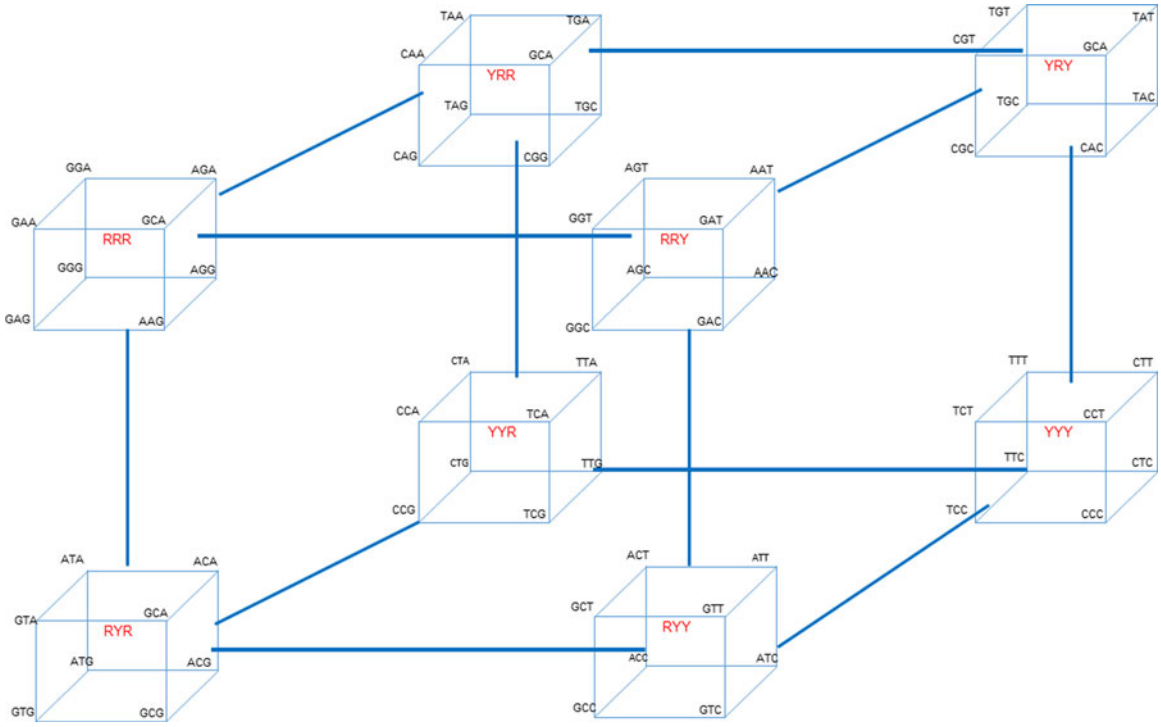


Fig. 2. The hypercube of all eight cube octet sequence classes (shown in red) located at each apex of the hypercube, illustrating the interconnectedness of the cycles associated with the full cascade of codon interconversions via sequential site changes. Transition mutations occur within a cube, while transversion mutations link one cube to another.

antiparallel, complementary codon. By the same criteria, a second group of significantly stable codons of the form, GCX, CGX, GGX, and CCX, also can be identified, although less resolved, where X is either T or A. It is noteworthy that collectively the GCX, CGX, GGX, and CCX families of stable codons code for Ala, Arg, Gly, and Pro, which are among the most abundant amino acids, and, save for Arg, also are considered ancient amino acids (Trifonov, 2000). Furthermore, when X = A or T, the complementary codons to this second group are XCG, XGC, XCC, and XGG. Except for Trp, these codons code for the amino acids Cys, Ser, and Thr, which, like Ala, Gly, and Pro noted above all are defined as ancient amino acids (Trifonov, 2000). This empirical correspondence between the stabilities of codons and the abundance as well as age of the amino acids for which they code raises the intriguing possibility of a stability-modulated, evolutionary shaping of the code.

These stable codon groups, and their corresponding, antiparallel codons, each occupy three positions within each of the eight cubes that make up the hypercube scaffolding that interconnect all 64 codons. This set of 24 'high-stability', codon groupings may have been energetically favored 'prebiotically'. As such, it is intriguing to note that all the other 40 codons can be generated/"evolved" from these most stable codons by, at most, three transition mutations; again suggestive of a stability-modulated, evolutionary shaping of the code.

To test the robustness of our conclusions, we conducted the same analyses for a dataset in which we reversed the polarities of the codon–complementary codon interactions to yield parallel codon couplets. This assessment yielded an altered energy spectrum, as well as changes in the stability rank order for the complementary codon couplets; particularly for the RRR/YYY and the YYR/RRY families of trimeric duplexes (columns 1 and 3 in Scheme 1). Further comparison of the antiparallel and parallel datasets also revealed differences in the energy changes associated with the sequential transition and transversion mutations. In the aggregate, these differential outcomes underscore the robustness of the correlations noted here between the stabilities of the antiparallel couplets and the shaping of the genetic code.

Correlations between larger domain DNA energy profiles and higher-order biological functions

One long-term goal is to define correlations between functional domains of the genome and the energy profiles of such domains. Some initial trends, that require further validation, include the suggestion by Klump and coworkers that protein-coding sequences predominately consist of codon domains of relatively uniform stability (Klump and Maeder, 1991). By contrast, Klump proposes that signal sequences exhibit less uniform and less stable domains, while also being more sensitive to local changes in cellular and sequence/structural environments, thereby allowing them to amplify a perturbation in a localized sequence. This biophysical behavior is what one would expect for a biological signal transducer. Coding sequences, by contrast, are less sensitive to environmental conditions, also consistent with their biological function to faithfully code for a protein. Much more research is required to establish a robust biophysical map of genomes to test such hypotheses. However, these intriguing early correlations should motivate such efforts, including a parallel analysis for RNA.

Concluding remarks

We have reviewed, presented, and integrated evidence that the iconic chemical genetic code also can be viewed as a differential

https://doi.org/10.1017/S0033583520000098 Published online by Cambridge University Press

energy code that influences biological outcomes. This perspective includes implications for differential, energy-based, molecular contributions to classic Darwinian evolutionary theory. In short, evolutionary pressures may well derive from the optimization of fundamental biophysical properties, as well as from the classic perspective of being driven to yield a functionally adaptive advantage for either a biopolymer or an organism.

Darwinian evolution, when proceeding over sufficiently long timeframes, leaves only the evolutionary 'winners' behind. This reality makes it difficult, if not impossible, to deduce, with any certitude, the precursors or evolutionary pathways that ultimately culminated in the current, evolved 'winners'. However, by evoking the laws of thermodynamics, it becomes possible via considerations of thermodynamic selection and linkages, as illustrated in the hypercube of Fig. 2, to speculate on what may have preceded these 'left behind winners'. It is precisely the beauty of thermodynamics that follows universal laws, under all conditions and times, that allows one to extrapolate backward from the 'left behind winners' to make informed speculations as to how these winners may have evolved from earlier remnants (molecular fossils).

Consistent with this perspective is our hypothesis that the evolution of the genetic code was shaped and modulated by the differential stabilities of complementary codons; a feature reflective of 'molecular Darwinism'. As thermodynamic fingerprints of such an evolutionary influence, we found correlations between the free energies of formation of antiparallel, complementary DNA trimers and their codon usage frequency. We also noted correlations between the stabilities of complementary codon couplets and those that code for 'ancient' amino acids. Collectively, our observations are consistent with a scenario in which the genetic code, driven by differential codon stabilities, evolved under the influence and regulation of a series of interlocking thermodynamic cycles. We proposed that these coupled energy cycles controlled the transition and transversion mutations of a group of the 24 most ancient ('prebiotic') and stable codon pairs, ultimately yielding the complete 64 codon code; via a form of 'thermodynamic selection'. As such, we suggested that the evolution of the genetic code exhibits contributions from both stability-driven 'molecular'/genotypic Darwinism as well as the more traditional, phenotypic Darwinism. As we stated in the Abstract, yet worth repeating, it is not surprising that evolution of the code was influenced by differential energetics, as thermodynamics is the most general and universal branch of science that operates over all time and length scales.

Going forward

Given the correlative examples noted here, going forward it seems justified to create a comprehensive energy map of the human genome; or for that matter, the genome of any organism of interest. The differential energy domains so characterized may correlate with known functionalities; or may reveal and yield insights into regions of yet defined function. Such profiling to create an 'energy genome' would yield a thermodynamic bridge between sequence, structure, and biological function.

Postscript: Shortly after the beginning of the 20th century, Albert Einstein (Schilpp and Einstein, 1949) declared:

> 'A theory is the more impressive the greater the simplicity of its premises, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression that classical thermodynamics made upon me. It is the only physical theory of universal content which I am convinced will never be overthrown, within the framework of applicability of its basic concepts'.

As biophysical chemists, the authors consider the ultimate exemplar/test of this assertion to be the demonstration that the molecular language and complexities of biology embedded in the genetic code can be rationalized in terms of fundamental thermodynamic principles.

Financial support. This research was supported by grants from the NIH GM23509, GM34469, and CA47995 (to K.J.B.) and NRF (Pretoria, RSA) grant GUN 61103 to H.H.K.

Conflict of interest. The authors declare no conflict of interest.

References

Aldana-González M, Cocho G, Larralde H and Martinez-Mekler G (2003) Translocation Properties of Primitive Molecular Machines and their Relevance to the Structure of the Genetic Code. *Journal of Theoretical Biology* **220**, 27–45.

Aldana M, Cázarez-Bush F, Cocho G and Martínez-Mekler G (1998) Primordial Synthesis Machines and the Origin of the Genetic Code. *Physica A: Statistical Mechanics and its Applications* **257**, 119–127.

Anderson S, Bankier AT Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Breitenberger CA and RajBhandary UL (1985) Some Highlights of Mitochondrial Research Based on Analyses of Neurospora crassa Mitochondrial DNA. *Trends in Biochemical Sciences* **10**, 478–483.

Breslauer KJ, Frank R, Blocker H and Marky LA (1986) Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3746–3750.

Crick FH (1968) The origin of the genetic code. *Journal of Molecular Biology* **38**, 367–379.

Darwin C (1859) *On the Origin of Species by Means of Natural Selection*. Neudr. Bruxelles 1969 Edn. London: Murray.

Delcourt SG and Blake RD (1991) Stacking energies in DNA. *The Journal of Biological Chemistry* **266**, 15160–15169.

de Ruiter A and Zagrovic B (2015) Absolute binding-free energies between standard RNA/DNA nucleobases and amino-acid sidechain analogs in different environments. *Nucleic Acids Research* **43**, 708–718.

Di Giulio M (2004) The Coevolution Theory of the Origin of the Genetic Code. *Physics of Life Reviews* **1**, 128–137.

Dill KA and Chan HS (1997) From levinthal to pathways to funnels. *Nature Structural Biology* **4**, 10–19.

Doktycz MJ, Goldstein RF, Paner TM, Gallo FJ, Benight AS *et al.* (1992) Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA. *Biopolymers* **32**, 849–864.

Eigen M (1976) *How is Information Formed? Principles of Self-Organization in Biology*. 80 Vol. Weinheim: Verlag Chemie, pp. 1059–1081.

Eigen M and Winkler-Oswatitsch R (1992) *Steps Towards Life: A Perspective on Evolution*. England: Oxford University Press.

Freeland SJ, Wu T and Keulmann N (2003) The case for an error minimizing standard genetic code. *Origins of Life and Evolution of the Biosphere: The Journal of the International Society for the Study of the Origin of Life* **33**, 457–477.

Gardin J, Yeasmin R, Yurovsky A, Cai Y, Futcher B *et al.* (2014) Measurement of average decoding rates of the 61 sense codons in vivo. *eLife* **3**, 10.7554/eLife.03735.

Greive SJ and von Hippel PH (2005) Thinking quantitatively about transcriptional regulation. *Nature Reviews. Molecular Cell Biology* **6**, 221–232.

Huguet JM, Bizarro CV, Forns N, Smith SB, Bustamante C, Ritort F *et al.* (2010) Single-molecule derivation of salt dependent base-pair free

https://doi.org/10.1017/S0033583520000098 Published online by Cambridge University Press

energies in DNA. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15431–15436.

Huguet JM, Ribezzi-Crivellari M, Bizarro CV , Ritort F*et al.* (2017) Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments. *Nucleic Acids Research* **45**, 12921–12931.

Klump HH and Maeder DL (1991) The thermodynamic basis of the genetic code. *Pure and Applied Chemistry* **63**, 1357–1366.

Koonin EV and Novozhilov AS (2009) Origin and evolution of the genetic code: The Universal Enigma. *IUBMB Life* **61**, 99–111.

Martinez-Mekler G, Aldana M, Cazarez-Bush F, Garcia-Pelayo R, Cocho G *et al.* (1999) Primitive molecular machine scenario for the origin of the three base codon composition. *Origins of Life and Evolution of the Biosphere: The Journal of the International Society for the Study of the Origin of Life* **29**, 203–214.

Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science (New York, N.Y.)* **117**, 528–529.

Miller SL and Urey HC (1959) Origin of life. *Science (New York, N.Y.)* **130**, 1622–1624.

Miller SL, Urey HC and Oro J (1976) Origin of organic compounds on the primitive earth and in meteorites. *Journal of Molecular Evolution* **9**, 59–72.

Novozhilov AS and Koonin EV (2009) Exceptional error minimization in putative primordial genetic codes. *Biology direct* **4**, 44.

Novozhilov AS, Wolf YI and Koonin EV (2007) Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biology Direct* **2**, 24.

Polyansky AA and Zagrovic B (2013) Evidence of direct complementary interactions between messenger RNAs and their cognate proteins. *Nucleic Acids Research* **41**, 8434–8443.

Porschke D and Eigen M (1971) Co-Operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic–oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *Journal of Molecular Biology* **62**, 361–381.

SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1460–1465.

SantaLucia J, Allawi HT and Seneviratne PA (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555–3562.

Schilpp PA and Einstein A (1949) *Albert Einstein, Philosopher-Scientist*. Evanston, Ill.: Library of Living Philosophers.

Trifonov EN (2000) Consensus Temporal Order of Amino Acids and Evolution of the Triplet Code. *Gene* **261**, 139–151.

Trifonov EN (2004) The triplet code from first principles. *Journal of Biomolecular Structure & Dynamics* **22**, 1–11.

Trifonov EN and Bettecken T (1997) Sequence fossils, triplet expansion, and reconstruction of earliest codons. *Gene* **205**, 1–6.

Wong JT (2005) Coevolution theory of the genetic code at age thirty. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **27**, 416–425.

Yarus M, Caporaso JG and Knight R (2005) Origins of the genetic code: the escaped triplet theory. *Annual Review of Biochemistry* **74**, 179–198.

https://doi.org/10.1017/S0033583520000098 Published online by Cambridge University Press