Published in final edited form as:

J Theor Biol. 2010 June 7; 264(3): 711–721. doi:10.1016/j.jtbi.2010.03.046.

The rules of variation: Amino acid exchange according to the rotating circular genetic code

Fernando Castro-Chavez^a

Universidad de Guadalajara, Centro Universitario de la Costa, Puerto Vallarta, Mexico

Abstract

General guidelines for the molecular basis of functional variation are presented while focused on the rotating circular genetic code and allowable exchanges that make it resistant to genetic diseases under normal conditions. The rules of variation, bioinformatics aids for preventive medicine, are: (1) same position in the four quadrants for hydrophobic codons, (2) same or contiguous position in two quadrants for synonymous or related codons, and (3) same quadrant for equivalent codons. To preserve protein function, amino acid exchange according to the first rule takes into account the positional homology of essential hydrophobic amino acids with every codon with a central uracil in the four quadrants, the second rule includes codons for identical, acidic, or their amidic amino acids present in two quadrants, and the third rule, the smaller, aromatic, stop codons, and basic amino acids, each in proximity within a 90 degree angle. I also define codifying genes and *palindromati*, *CTCGTGCCGAATTCGGCACGAG*.

Keywords

Codons; Proteins; Isoforms; Polymorphism; Replacement; Trinucleotides; Triplets

1. Introduction

It is rightfully assumed that the genetic code was complete from the very beginning; indeed, it was the condition for the origin of life and its variability, being valid for every living organism (Rakocevic, 2004).

What motivates my efforts for defining the rules of biological variation is that the genetic code was theoretically predicted (Gamow, 1954), even if the theory was not completely perfect at the beginning (Hayes, 1998), and then the genetic code was found ([Martin et al., 1962], [Khorana, 1965] and [Legato, 2010]). The seminal work of Mendel on the genes (that he called “factors”) was based on his statistical study of phenotypes (Mendel, 1865), published in an obscure journal, being fully ignored at first, then discovered and recognized ~35 years later (Bateson, 1902).

This work focuses on the forces that influence the patterns of naturally occurring variation, and how those patterns can be explored to understand human disease, which mathematically and statistically can be seen as the outlier of the normal series.

The classic circular genetic code representation originally drawn by Bresch and Hausmann (1972) is the circular genetic code we consider in this article and has been used successfully

^aPresent address: Department of Medicine, Atherosclerosis and Vascular Medicine Section, Methodist DeBakey Heart Center, Baylor College of Medicine, Houston, TX, USA. Tel.: +1 713 798 4177; fax: +1 713 798 4121. URL: <http://fdocc.blogspot.com/2010/04/gencode.html>

for the balanced computation of amino acid nucleons by ShCherbak (2008) and is also useful for educational purposes, as stated by Miller and Levine (2007), who found that students learned how to use the circular genetic code in a matter of minutes as it was simpler to learn, easier to use, and it encouraged them to investigate genes and gene sequences in greater depth, resulting in a lesser hesitation to use it, and feeling more at ease in using the circular genetic code to explore the process of translation as opposed to the traditional genetic code table, which is rectangular, with the first, second, and third letters of the code occupying different axis on the rectangle deemed to be tremendously confusing. Jimenez-Montano (2009) noticed “a very singular situation in science history: the inseparability of physical–chemical, informational (communication) and biological aspects in the description of the genetic code”, Sanchez and Grau (2005) considered “the genetic information system as a communication system” while Barbieri (2008) concluded that “every living cell contains all four components of semiosis (signs, meanings, code, and ‘code-worker’ [in *italics* my words]).”

Limitless differences among individuals (polymorphisms) allow the proteins of mammalian systems to normally perform their multiple functions. The key to variation that preserves the protein function is retention of specific nucleotides in the DNA codon that encode amino acids with similar polarity or hydrophobicity (conservative, non-harmful, or fault-tolerant substitutions).

Variations due to adaptations improve the survival of an individual, i.e., homologous proteins for organisms under diverse environments exhibit identical folding patterns and shapes but different amino acid sequences, preserving very few key amino acids. There are multiple versions for the same protein, both in different organisms and in different tissues of the same organism.

There are variable regions within a protein, some of them unimportant for its function, i.e., the first residues of some equally functional proteins, when compared, not only change in the type of amino acids present, but also in the number of residues.

The extremes of variability are the proteins whose sequences are highly conserved (ubiquitin, histones H3 and H4, ribosomal proteins, calmodulin, etc.), and in the other extreme the non-conserved proteins, the ones not subject to any constraint, i.e., the non-functional flanking N- and C-terminal segments of the HMG box of SRY (the major sex-determining locus), and the fibrinopeptides with a length of 20 amino acids and thought to be functionless (Strachan and Read, 1999).

In some cases, only a few amino acids in key positions within a protein will be present in all organisms, i.e., in one study comparing 226 globin sequences, only two histidines were found to be conserved, one binds and releases oxygen (Karp, 2008) while the other is linked to the heme iron (Bashford et al., 1987). The most striking example described by Karp was the amino acid replacement of halophilic micro-organisms that enabled functioning at extremely high salt concentrations; the surface of the halophilic isoform of malate dehydrogenase, unlike its non-halophilic counterparts, is coated with glutamic and aspartic acid, whose carboxyl groups are capable of competing against the salt to capture water molecules (Dym et al., 1995). Di Giulio (2005) found a correlation between barophily (the resistance to pressure) and a higher abundance of small amino acids when compared to non-barophilic microorganisms. Some proteins of micro-organisms lack big amino acids, maybe due to space constraints, e.g., the ferredoxins of *Clostridium butyricum*, in which Arg, Lys, Leu, Met, His, Tyr, and Trp are missing (Yockey, 2005). A protein that lacks one of the 20 common amino acids is called a protein lipogram (White III and Dhurjati, 2006).

In my previous research, it was possible to demonstrate the coordinated change of gene expression for proteins closely associated metabolically; the Krebs cycle, the respiratory chain and the beta-oxidation pathways were up-regulated, while the genes producing proteins for the fat unsaturation and cholesterol synthesis were down-regulated ([Castro-Chavez et al., 2003] and [Castro-Chavez, 2004]). These findings, plus a set of coordinated changes recently discovered in atherosclerosis (Vickers et al., 2009) stimulated my current research on the functional variation within proteins.

The third position of the codons is highly reiterated (redundant) and according to current evaluations allows ~67% of synonymous substitutions, the first base position is less reiterated and allows only ~4% of synonymous substitutions, while the second base allows none (Strachan and Read, 1999). These reiterations in the genetic code minimize possible damages to the function of the protein. Siemion and Siemion (1994) found that the amino acids with *R* (purine, pur) in the third position are of higher structural importance than the amino acids coded with *Y* (pyrimidine, pyr), and that the amino acids that prefer alpha-helical over the beta-sheet conformation favor the codons with *R* in their third position, and those that favor the beta-sheet conformation favor the codons with *Y* in the third position.

Since the earlier days of its discovery, the order in the genetic code was recognized. Woese (1965) declared that the correlation between codons assigned to related amino acid “is not only significant but very extensive”. The most recent review on the nature of the genetic code emphasizes that the genetic code can be seen as optimized for the tolerance of transitions (purines (*A*, *G*) replaced by purines, or pyrimidines (*T*, *C*) by pyrimidines), which are the most significant nucleotide substitutions (Jestin and Kempf, 2009). The error-minimizing property of the genetic code has a significant influence in differentiating synonymous codon usage, for example between thermophilic and mesophilic prokaryotes (Basak et al., 2007), and its capability to enhance variation within proteins (Zhu and Freeland, 2006). However, the existing literature has not adequately addressed the importance of the circular genetic code as a source to predict variation based on its rotational periodicity; here, this issue is presented in a novel way and thought to be useful for the early detection of harmful mutations to minimize and/or to prevent its later consequences.

The main reason why I use the classic circular genetic code as a better representation than the ordinary rectangular code is its rotational ability and resonance, defining resonance as the natural ability of the circular genetic code to reflect or mirror on similar amino acids on rotation. As we will see, if we rotate its quadrants by 90°, 180°, or 270°, we will always have every hydrophobic amino acid codon (with *U* at the center, which is *T* in the DNA) at the same relative position or resonance. The same can be said of the acidic amino acids and their amides, and of synonymous codons positioned in different quadrants for amino acids that have six codons (Leu and Arg). In addition, rotations of less than 90° can be achieved within the same quadrant, like the Ser/Thr non-harmful replacements. Thus, the rules advanced here require resonant rotations of non-harmful amino acid replacements to be done either in 4, 2, or 1 quadrant, respectively, for rules 1–3. The rotational capacity of the circular genetic code described here makes it a strong candidate in computational biology or bioinformatics.

The focus of the present analysis is the positional relation of codons within the quadrants of the circular genetic code to infer their common properties, both to predict compatible amino acid exchanges as well as their pathological substitutions. Among other genetic code representations, I found the quasi-sphere (Yang, 2004; the third figure in Hayes, 2004), the cube ([Jimenez-Montano et al., 1996] and [Karasev and Stefanov, 2001]), the octets ([Wittmann, 1963] and [Rumer, 1966]), the clusters (Di Giulio, 1989a), tetrahedrons and rectangles (Demongeot et al., 2006), including other two-dimensional non-circular versions

([Yu, 2007], [Khrennikova and Kozyrev, 2007] and [Jimenez-Montano, 2009]; etc.) However, the focus of this article is on the circular bi-dimensional genetic code to be used as the best representation for a reproducible tool to predict and prevent hereditary diseases.

I provide the functional definition of a codifying gene as “the set of trinucleotides that make a protein”. A less practical definition, taken from the point of view of the genomic *loci* has been proposed by Gerstein et al. (2007): “Gene: A union of genomic sequences encoding a coherent set of potentially overlapping functional products”. However, in my functional definition of a gene from the point of reference of proteins, different related proteins will have different related genes even if they originated in the same genomic locus. Using the same functional logic, Kane (2004) defines a gene as “both the static chemical compound and the dynamic template executed through the genetic code”; she also provides highly related definitions for the genetic code: “the set of invariant relationships between DNA and protein”, “the universal dictionary for the conversion of DNA to protein”, and “the molecular dictionary of the cell”, while, according to Voie (2006), the genetic code is a time-independent language to build functional biomolecules. By replacing one amino acid with another, most frequently the protein function is not affected (Yanofsky et al., 1966). Goldberg and Wittes (1966) in characterizing amino acid exchange within functional proteins noticed that: (1) except Tyr, hydrophobic amino acids were able to be exchanged to others of their group with 100% frequency, (2) the “most-protected” group of amino acids (Ala, Gly, Leu, Pro, Arg) with an average frequency of 7.7% each (according to studies done at the time in micro-organisms) are six or four times reiterated, and have a maximum GC, while the “least-protected” ones (Met, Tyr, Phe, Lys, Ile), having an average frequency of 3.7% each, are twice reiterated and lack GC, (3) the residues best protected are the preponderant ones in proteins with the exception of Pro and Arg; the unique role of Pro in polypeptide conformations provides an explanation for its added protection and, in the basic group, the apparent overprotection of Arg compensates for the underprotection of Lys, located in a quadrant's distance of each Arg set of codons (within a 90 degree angle), (4) residues Cys and His, both of which appear to be of special importance in enzyme structure, do not fall in the “least-protected” class, both having codons containing 50% of GC, (5) the exclusion of transversions decreases even more “the frequency with which a hydrophilic amino acid mutates to the hydrophobic type”, (6) amino acids “appear to mutate to codons of the same group far more frequently than expected from random assignment of codons”, and (7) triplets high in GC content have “a greater pairing specificity and greater resistance to mutation”, concluding that the genetic code “seems to be organized so that the exchange of hydrophobic and hydrophilic amino acids as a result of one-step mutations is minimum”, they concluded.

Alff-Steinberger (1969) posed the question, “is the genetic code an error-minimizing code?”, discovering by computational methods that “for the natural code, single-base substitution in the first position of the codon tends to result in the substitution of an amino acid more similar to the original amino acid than would be expected from a random code”, and that the second position of the codon is the one playing the largest role in determining the properties of the amino acids.

A complex drawing of all possible codon replacements can be seen in the first figure by both Maeshiro and Kimura (1998) and Hayes (2004), who noticed a growing recognition that the genetic code may encompass more information than just the mapping from codons to amino acids, as synonymous codons may not always be completely equivalent; codon frequencies are not random or uniform among the codons that specify a given amino acid, as some may be common and some rare, and these usage biases can vary both within and between genomes, helping to regulate the rate of protein synthesis; “if the transfer RNA that matches a codon is rare, then transcription of genes including that codon will be slowed. For some

proteins there is evidence that such pace-setting codons help ensure correct folding of the amino acid chain". Hayes also wrote that the genetic code seems to be organized so that common substitutions cause little change in polarity or hydrophobicity. Using a related bootstrap criterion (the observed codons), Freeland et al. (2000) noticed that the genetic code "appears at or very close to a global optimum for error minimization: the best of all possible codes". Maeshiro and Kimura (1998) calculated the robustness and changeability of genetic codes, separating Ser in two nodes, while, better of, Hayes (2004) located Ser in the same node of the genetic code.

2. Materials and methods

For the investigation of genetic variation, it is helpful to use the circular genetic code as illustrated by Koolman and Roehm, 2005 [J. Koolman and K.H. Roehm, *Thieme Color Atlas of Biochemistry* (2nd ed), Thieme, Stuttgart (2005) (p. 249). Also see their 1996, 1st ed], based on the original work done by Bresch and Hausmann (1972); more recently, Miller and Levine (2007) used it for didactic purposes. The circular genetic code is composed of four quadrants, with the first nucleotide of the codons at the center (*U, C, A, G*)¹; then the second nucleotide letter in the second layer (layer composed of 4×4 letters), and the third nucleotide letter in the third and last layer (layer with 4×4×4 letters). The next rules of variation are presented from the point of view of the circular genetic code and are in no way exhaustive, the fine-tuning on the function of amino acid sequence positions within proteins is beyond the scope of this study.

3. Results

3.1. Relationship between codon base change and amino acid polarity or hydrophobicity

The rules that allow amino acids to be interchangeable specify that a nucleotide change within a codon, causes a change in the corresponding amino acid within a peptide or protein, a change that under normal circumstances preserves the full function of a protein. [Fig. 1], [Fig. 2], [Fig. 3] and [Fig. 4] depict the rules of variation based on the positional relation of the codons for amino acids within the circular genetic code.

Rule 1—There is a positional homology or “*resonance*” for the hydrophobic amino acids in the four quadrants of the circular genetic code and that every codon with a *U* at the center (*NUN* for RNA, which is *NTN* for DNA, N being any nucleotide) encodes only for essential hydrophobic amino acids (Fig. 1). The hydrophobic and essential amino acids can be considered here as the “engines of variation” constituting the collective “clusters of hydrophobicity” (Margoliash and Smith, 1965). Examples of such exchanges are Ala to Val (Lecomte et al., 1999), Val to Ala (Worn, 1998), Ile to Val (Wang et al., 2009), Val to Ile (Hayes et al., 2000), Val to Leu (Shemirani and Muszbek, 2004), Leu to Val (Junca et al., 2002), Leu to Ile (Elahi et al., 2003), Ile to Leu (Ieiri et al., 2000), Val to Phe (Koene et al., 1997), Leu to Met (Lu and Elzinga, 1977), Phe to Ala or to Ile (Lutsenko et al., 2007), Trp to Ala (Ueda et al., 2009), Phe to Leu (Hoffmeyer et al., 2000), Leu to Phe (Jann et al., 2008), etc. Notice that this rule does not say that all hydrophobic codons are located in the same quadrants' position.

¹Abbreviations: **Amino acids**: A: alanine (Ala), V: valine (Val), I: isoleucine (Ile), L: leucine (Leu), M: methionine (Met), F: phenylalanine (Phe), W: tryptophan (Trp), D: aspartic acid (Asp), N: asparagine (Asn), E: glutamic acid (Glu), Q: glutamine (Gln), R: arginine (Arg), K: lysine (Lys), S: serine (Ser), T: threonine (Thr), G: glycine (Gly), P: proline (Pro), H: histidine (His), C: cysteine (Cys), Y: tyrosine (Tyr). **Nucleotides**: *U* or *u*: uracil, *C*: cytosine or *c*, *A*: adenine or *a*, *G*: guanine or *g*, *T* or *t*: thymine, *Y*: pyr, pyrimidine; *R*: pur, purine; *N*: aNy nucleotide.

Rule 2—There is a positional homology in two of the four quadrants of the circular genetic code for equal (synonymous codons) or related amino acids (Fig. 2). Examples of such exchanges are Asp–Asn (Bellapadrona et al., 2007), Asn–Gln or Asn–Asp (Feller et al., 1997), Asp–Gln (Smith et al., 2001), Asn–Glu (Gomez-Casado et al., 1996), Glu–Asn (Perria et al., 2006), Thr–Leu (Lang et al., 2004), Gln–Asn (Beyermann et al., 1996), Glu–Asp (Chen and Tsien, 1997), Glu–Gln (Luo and Horowitz, 1993), etc. Notice that this rule does not say that every resonant or contiguous codon in two quadrants is exchangeable.

Rule 3—There are equivalent amino acids or functional codons (i.e., stop codons) in the same quadrant (Fig. 3). Examples of such exchanges are Glu–Asp (Fan et al., 2007), Asp–Glu (Bag et al., 2007), Ser–Thr (Oppermann et al., 1997), Thr–Ser (Blikstad et al., 2008), Ser–Tyr (Sato and Kuroda, 2001), Gly–Ala (Plaseska et al., 1990), Ala–Gly (Linsel-Nitschke et al., 2006), Phe–Tyr (Viguera et al., 2002), Tyr–Phe (Abe et al., 2008), Tyr–Leu (Wang et al., 2002), Phe–Trp (Tradler et al., 1997), Trp–Phe (Vogtherr et al., 2002), Trp–Tyr (Asensio et al., 2000), Tyr–Trp (Steinberg, 2008), etc. Notice that this rule does not say that every codon within a quadrant is exchangeable.

4. Discussion

There have been other attempts to present the amino acids in a circular “gray code”. For example, improving on (Swanson, 1984a) and (Swanson, 1984b), Bosnacki et al. (2003) placed the amino acids in four groups, writing that perhaps the main hypothesis suggested by Swanson was that the genetic code was optimized to be robust with regard to translation errors “rather than random mutations”. However, Swanson (1984a) defined a gray code as “an error-resistant binary encoding scheme, which is also a mathematical expression of the principle of minimum change coding.” Nucleotide matrices have been used by Negadi (2001), and by Crowder and Li (2010), the latter ones summarizing the current status of their studies by declaring, “although our study does not have direct applications to biological problems yet, it is our hope that the matrix results will help organize, store, and retrieve data in a systematic way so that hidden patterns in genetic sequences can be recognized effectively by computers or humans.” Di Giulio (1989a), based on Woese et al. (1966), obtained four clusters of amino acids based on the nucleotide in its second codon position, on polarity values and on absolute values for standardized molecular volumes. The same year, Di Giulio (1989b) published, that amino acids exhibiting similar physical properties tend to have similar codons.

However, none of these gray codes, matrices or clusters proved useful for predictability or rotational periodicity, which is available only in the “resonant” and rotating circular genetic code.

The circular genetic code is a different concept than the “circular code”, which is a given set of codons able to be read in only one frame without a designated initiation signal (Arques and Michel, 1997), developed into a proposed tool to identify the proper reading frame of genes when the start codon is missing (Arques et al., 2002).

According to Berleant et al. (2009), the standard rectangular codon table, although a brilliant advance, ignores important properties of genetic coding, including that: (1) similar codons tend to code for physico-chemically similar amino acids, (2) if position 2 of a codon is *U*, the codon is for a hydrophobic amino acid, (3) heavier amino acids tend to have fewer codons, and (4) frequently appearing amino acids tend to have more codons. They feel that an ideal representation of the genetic code continues being elusive today. They state that an important aspect of the true genetic code is its ability to tolerate mutations and mistranslations. However, the results of analyzing the circular genetic code described here

(which to Berleant is a “*codon wheel*”), without exceeding in complexity, are capable of overcoming the problems found by Berleant in the standard codon table: (1) the amino acid similarity can be tracked by the positional rotation of the three rules of variation presented here, (2) the *U* at the center of hydrophobic amino acids is here rule number one, (3) heavier or aromatic amino acids fall within quadrant 1, with rule number three applying to them, and (4) amino acids with more than four codons, either frequent or infrequent within proteins, have positional resonance according to rule number two.

To find how the circular genetic code exhibits the amino acids that can be interchanged, we can rotate quadrant one by 90° at a time to see the correspondence of the hydrophobic codons with *U* at the center in every quadrant (Rule 1), as well as contiguous correspondence of the codons for the amides Gln and Asn in quadrants two and three, respectively, with their corresponding acidic amino acids Glu and Asp perfectly matching them in quadrant four, as well as the location of codons for Arg and Leu in two different quadrants (Rule 2).

The third nucleotide in the codons is known for its synonymous replacements to reduce damage by harmful mutations. The second nucleotide of the codons is the core of stability; for example, U at the center is present only in codons for the hydrophobic essential amino acids: Leu, Val, Ile, Phe, and Met. The six codons for the hydrophobic Leu, the four codons for Val, the three codons for Ile, the two codons for Phe, and the single codon for Met, all of them have U in their center. The codons for the acidic amino acids (Glu, Asp) and their amides (Gln, Asn) have an A in the center. Aromatic amino acids and stop codons are grouped in the first quadrant while the small and the acidic amino acids are present in the fourth quadrant.

Goldberg and Wittes (1966) described some of the most notable correlations for the study of amino acid exchange, suggesting that protective mechanisms could be acting at two levels: (1) nucleotides, where a high *GC* content may be reducing “the rate of mutations and errors in protein synthesis” and (2) the organizational level, where the effects of a base change are minimized by the reiteration (the so-called “redundancy”) of codons, “and by the connectedness of codons for similar amino acids” (their words). Therefore, they found that the “best protected” amino acids are those with maximum *GC* content and codon reiteration, indicating that the organization of the genetic code reflects the composition of cellular proteins by: (1) groups of chemically similar amino acids that have similar codons and (2) the degree of protection afforded a group, as measured by the number of codons allotted to it, which “is proportional to the frequency of appearance of its amino acids in proteins.”

Pelc (1965) described the correlation between codons and its amino acids and found that the codons with nucleotide *U* at their center corresponded to hydrophobic amino acids (Phe, Leu, Ile, Met, Val), while the ones with *C* at the center were related to weakly polar amino acids (Ser, Thr, Pro, Ala), and those containing an *A* (Asp, Glu, Asn, Gln, His, Lys, Tyr) or a *G* (Arg, Cys, Trp, Ser, Gly) at the center included the strongly polar amino acids. Alff-Steinberger (1969) noted that, “the second position of the codon carries the most information, i.e., plays the largest role in determining the properties of the amino acid.”

In one of Crick's last reviews of the genetic code (Crick, 1968), he published the rectangular genetic code (also in the order *U, C, A, G*, from left to right and top to bottom), recognizing that “even allowing for the grouping of codons into sets, the amino acids do not seem to be allocated in a totally random way. For example, all codons with *U* in the second place code for hydrophobic amino acids. The basic and acidic amino acids are all grouped near together towards the bottom right-hand side. Phenylalanine, tyrosine, and tryptophan all have codons starting with *U*, and so on. It is very difficult not to imagine regularities in even a random

grouping but nevertheless the general impression is that “related” amino acids have to some extent related codons (Epstein, 1966)” [Crick's emphasis and reference]. Epstein (1967), using also the rectangular genetic code, pursued his observations on the non-randomness of amino acid exchanges in homologous proteins. Lagerkvist (1978) observed that if the first two nucleotides in the codon are *G* and *C*, which are the “strong” codons, they always represent full four-codon families (*GGN*-Gly, *CGN*-Arg, *GCN*-Ala, *CCN*-Pro), while “weak” codons (*A* and *U* as the first two nucleotides), never do. He also indicated that “mixed” codons [(*Y*) pyr/(*R*) pur] at the right side of the traditional rectangular genetic code never include “codon families”; whereas, mixed codons at the left side always constitutes a codon family. He saw that codon families have a much higher probability to appear in the left side of the rectangular representation. Di Giulio (1989a) wrote that “the positions that are more protected, in terms of the effects of mutation, seem to be those of the amino acids coded by codons with *U* in the second position”, and those are precisely the ones that integrate the first rule of variation described here.

The complex tables drawn by Biro et al. (2003) emphasized the importance of the central nucleotide in the codons, anticipating that purines control the charge while pyrimidines determine the amino acids polarity. In relation to the atomic structure of the amino acids, Davydov (1998) found that all oxygen or nitrogen ending amino acids have codon doublets that include the nucleotide *A* and that all solely carbon ending (excluding Ala), and sulfur ending amino acids have codon doublets that include the nucleotide *U*, with proline and glycine remaining out of the classification due to their peculiar structure. By analyzing the rectangular genetic code, Qiu and Zhu (2000) concluded that, “the obvious orderliness within the genetic code and within the codons themselves must have not been the result of chance, but the reflection of the history of the genetic code”. However, neither of the notable scientists mentioned thus far gave note to the cyclic (rotational) periodicity and predictability of the genetic code, something that was harder to do on the basis of a rectangular genetic code, easier to note using a circular representation.

As demonstrated here, the classic circular genetic code is a better tool in the study of the amino acid order and reproducibility within proteins and it is suggested as the basis for a computational program to detect functional variability.

The six codons for the basic amino acid Arg, both in quadrants 2 and 3, have a **G** at their center; however, the codons for Ser have at their center either one of the “strong” or *three hydrogen bond* nucleotides: **C** for its four codons present in quadrant 1 and **G** for its two codons present in quadrant 3. The four codons for Serine's related amino acid Thr also have a **C** at their center in quadrant 3. The smaller amino acids share *three hydrogen bond* nucleotides at the core of their codons: a **G** for Gly and a **C** for Ala. Ala and Trp (both hydrophobic, the small and the big), have **C** and **G**, respectively, at the center of their codons. The only two codons for the basic amino acid Lys have an **A** at their center. Two stop codons have an **A** at their center while the remaining one has a **G**. Appendix A presents a list of codons, their properties, and exchange. Only Met (start codon) and Trp lack reiterations.

As for function, the start codon (*AUG*), encoding for Met, is in the opposite quadrant to that of the three stop codons. Also, the amino acids that have odd codons are present in opposite quadrants (three and one, respectively), i.e., the products of a single codon (both hydrophobic and essential: Met and Trp), or three codons (essential amino acid Ile and the stop codons). The presence of the stop codons in the genes is strictly non-randomly controlled; if it were random, we should be expecting some 30 stop codons per mRNA, when only one stop codon is present and necessary (Strachan and Read, 1999). According to the “Codon Usage Database” (Nakamura et al., 2000), the three most used codons in humans

are *CUG* (Leu), *GAG* (Glu), and *CAG* (Gln), while the less used codons are the stop codons: *UGA*, *UAA*, and *UAG*.

In a previous analysis of the molecular basis of variation, I discovered a methodological artifact that I called *palindromati* (Castro-Chavez, 2005), whose sequence I reported earlier —*CTCGTGCCGAATTCGGCACGAG* (Castro-Chavez, 2004). This sequence or fragments of the sequence contaminate many other sequences in the *GenBank*. I define *palindromati* as a laboratory artifact produced by a biological host interacting with a vector and generating full or partial palindromes or palindromic sequences capable of self-annealing, preventing them from being degraded by restriction enzymes (e.g., *EcoRI*) while binding in one heterogeneous strand (chimera) two independent nucleic acid sequences. “A palindromic sequence is a sequence that reads the same on the complementary strand” (Nikolajewa, 2006), an example of a *palindromati* can be seen at Li et al. (1999). Based on those results, we need to evaluate carefully some amino acid sequences as possible methodological artifacts in our sequence comparisons, for example (in their one letter abbreviation), NSARARA, NSARA, ARARA, RAEF, SCRIRHE, GTSSC, etc., as well as the non-relatedness of the sequences that appear before and after the artifact linker (Castro-Chavez, 2005).

Examples of some non-conservative changes that break the rules of variation and may end with diseases are Gly–Arg (Armstrong et al., 1996), Pro–Leu and Ala–Glu (Shin et al., 2000), Gly–Asp, causing cystic fibrosis (Hamosh et al., 1992), Gly–Val (Chiu et al., 1993), Glu–Val, causing sickle cell anemia (Serjeant and Serjeant, 1972), Val–Phe (Stafforini et al., 1996), Arg–Gln (Larsen et al., 1998), Glu–Ala (Sluyter et al., 2004), Ser–Cys (Chen et al., 2001), Ser–Pro (Chen et al., 1992), Arg–Pro (Sheppard et al., 1993), Pro–His (Humphries et al., 1990), His–Pro (Dalla-Venezia et al., 1993), His–Tyr (Dear et al., 2004), Tyr–Ala (Lucio and Mazzoni, 2008), Leu–Pro (Belsham et al., 1995), Val–Cys (Saito et al., 2000), and Phe–Cys (Mundo et al., 2001).

By regrouping the nucleotides of the codons within a circular genetic code, in the non-conventional order 2–1–3, starting at the center (clockwise: *C*, *G*, *A*, *U*), Pohlmeier (2008) was able to bind the six codons for Ser into one cluster, the six codons for Arg into another cluster, the six codons for Leu into another cluster, and the three stop codons into another cluster, as well as the acidic amino acids with their respective amides surrounding them in one macro-cluster. Every codon having *U* at its center was also macro-clustered in the fourth quadrant. However, the cyclic positional relation between quadrants as well as the quadrants' boundaries (except for quadrant four) was completely lost when compared with the classic circular genetic code described here.

When Davis (2009) brought the work of Pohlmeier to the in-frame order of 1–2–3 for the nucleotides, keeping at position two only the four bases, (clockwise: *A*, *G*, *C*, *U*), the only quadrant that remained intact was the fourth, with all the codons with U at the center macro-clustered together (with Leu remaining at the same position as in Pohlmeier's), while the individual clusters for Ser, Arg, and the stop codons were positioned counterclockwise. In Davis, the macro-cluster of acidic amino acids with their amides was lost, as well as their positional resonance according to the classic circular genetic code. In his most complex representation, Davis (2009) added the tRNA factor, as did Filatov (2009) in his rectangular table based on the division of amino acids into two aminoacyl-tRNA-synthetase classes, ordered by their molecular masses and the coding bases. However, in order to do a simplified and direct comparison of sequences using the circular genetic code as point of reference, the codons and their resulting amino acids are sufficient. The classic circular genetic code has the advantages of being cyclic, periodic, and predictable, adequately suited

for the early detection of harmful mutations and the possibility of reducing to a minimum their consequences in a timely fashion.

When I transformed the novel binary genetic code that uses 0 and 1 (pyrimidines: *U*, *C*=0; purines: *G*, *A*=1) by Nikolajewa et al. (2006) into a circular genetic code, I obtained a rotational resonance for the hydrophobic amino acids only within the first half of the circle, and not in the whole circle as happens in the circular genetic code. Nikolajewa was able to locate Ser in the same relative position in three of their octants (000-2, 001-2, and 110-2), similarly the acidic amino acids (110-3, 111-3) and the amide Gln (011-3), except for Asn (110-4); however, a confounding ambiguity was introduced by allocating Ile and Met at the same position of the 101 series (the fourth slot, or 101-4). The same happened to the Trp and one of the stop codons, which were lumped together in the same position of their 110 series (the second slot, or 110-2). Hence the classic circular genetic code remained as the best representation for studies of amino acid exchange periodicity and predictability.

The circular genetic code drawn by Moura et al. (2010) includes the codons for the human mitochondrial genes. In human mitochondria, only 13 of the ~900 proteins of its proteome are encoded by its genome (Eltner et al., 2008). Ambrogelly et al. (2007) drew the 16 alternatives found in mitochondria from diverse organisms in the circular genetic code and declared that the 10 codon reassignments in prokaryotic and eukaryotic nuclear codes are a subset of the 16 mitochondrial ones; that is, every single nuclear codon alteration has been observed in mitochondria as well. Thus, the same changes seem to occur through a conserved codon reassignment process for both prokaryotic and eukaryotic nuclear and mitochondrial codes, which seem to be a function of the size of its host. However, the universal genetic code is used without changes in chloroplasts and in mitochondria of green plants (Jukes and Osawa, 1990).

This research is not exhaustive of all the possibilities of variation within similar functional proteins. However, it advances the possibility for a stronger predictability of non-harmful variations, which is helpful for defining thousands of unclear changes under the current knowledge and databases. For example, in *VarySysDB*, the database recently developed by Shimada et al. (2009), the next changes remain unclear, while by using the rules of variation reported here, we may conclude that in general, these variations were non-harmful: (1) Agrin (AGRN): *Gtc/Atc* bringing a Val/Ile change (Rule 1), (2) protein kinase C zeta (PRKCZ): *ttC/ttG* causing a Phe/Leu replacement (Rule 1), (3) Von Willebrand factor A domain-related protein isoform 1: *gCg/gTg* changing Ala/Val (Rule 3), (4) multiple epidermal growth factor-like domain 6 precursor (MEGF6): *gCg/gTg* causing Ala/Val change (Rule 3), and (5) stromal cell-derived factor 4 (SDF-4): *Cag/Gag* changing Gln/Glu (Rule 2).

On the other hand, many unexpected changes predicted by the rules reported here are empirically accounted for by the same *VarySysDB* database as non-harmful; for example: (1) nucleolar complex protein 2 homolog (NOC2L): *tCg/tTg* not a harmful Ser/Leu replacement (Rule 2), *cGt/cAt* not harmful Arg/His (Rule 3), (2) 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase eta-2 (PLC-eta2): *Cgg/Tgg* not harmful Arg/Trp (Rule 2), (3) taste receptor type 1 member 1 precursor (TAS1R1): *tGc/tAc* not harmful Cys/Tyr (Rule 3), *Cgc/Tgc* not harmful Arg/Cys (Rule 2), (4) kinesin-like protein KIF1B (Klp): *tAt/tGt* not harmful Tyr/Cys (Rule 3), *Aaa/Gaa* not harmful Lys/Glu (Rule 2), *aCc/aTc* not harmful Thr/Ile (Rule 3), (5) Von Willebrand factor A domain-related protein isoform 1: *Gcc/Ccc* not harmful Ala/Pro (Rule 2), etc. Notice that the latter non-harmful amino acid exchanges are specific for the mentioned proteins; according to their position, they may be of more consequences for another protein, as exemplified by the next references: Ala-Pro ([Armstrong et al., 1996] and [Wang, 1996]), Arg-Cys (Lorenzo et al., 1993), Arg-His

(Ridgway et al., 1997), Glu–Lys (Bunn et al., 1982), Cys–Tyr (Shin et al., 2000), etc. These observations account for further study of the amino acid position within proteins and its effect on functional variation as well as restrictions to these rules based on positional grounds, also needed to define a protein's basic 3-D structure.

The rules of variation described here can be used to assess any study related to human polymorphisms such as the PGP (Church, 2005), the 1000 Genomes Project (SoRelle, 2008; Via et al., 2010), multiple cohort studies like ARIC, Oregon-SUDS, CHS, etc. (Arking et al., 2010), and for studies related to any other organism, for example in plants (Ruiz-Rojas et al., 2010), cattle (Snelling et al., 2010), birds (van Bers et al., 2010), bacteria (Huang et al., 2008), etc.

5. Conclusion

The rules of variation presented here are: (1) same position in the four quadrants for hydrophobic codons, (2) same or contiguous position in two quadrants for synonymous codons (codifying for the same amino acid) or related codons, and (3) same quadrant for equivalent codons. These rules can be the basis of a computational program with the genetic code at its core to compare genetic sequences and proteins to determine their range of non-harmful variation and deleterious mutations causing hereditary diseases. When the rules of amino acid variation are broken, we may have a harmful mutation causing disease at key positions by the next reciprocal exchange (1) when hydrophobic amino acids are substituted by polar amino acids and (2) when histidine, cysteine, and proline are replaced by any other amino acid.

While hydrophobic amino acids seem to have high plasticity of replacement among them, there is also exchange within basic, aromatic, and acidic amino acids and their amides. The special amino acids proline, cysteine, and histidine seem to be more limited in their exchange and are located in a quadrant's distance of each other (within a 90 degree angle).

The reiterations, equivalences, and resonances in the circular genetic code minimize possible damages due to forbidden nucleotide exchanges (the statistical outliers). These rules underscore the robustness of the circular genetic code and its resistance to harmful mutations while promoting physiological and phenotypic exchanges under normal conditions. The classic circular genetic code has the great advantage of its periodicity, which is what makes it predictable. Additional fine-tuning of these rules based on other parameters like amino acid position, function, modification, or environment will greatly improve our understanding of genetic variation, the frequency of these changes within populations, and the early detection of pathological changes. We can conclude that the genetic code is an intelligent design that maximizes variation while minimizing harmful mutations.

Acknowledgments

Thanks to Joel D. Morrisett for reviewing an earlier version of this work, to Estefano Macedo-Reyes, Gerd Brunner, Patrick and Edward Snyder-Castro, for their valuable suggestions, to Christie Ballantyne, Diana Hunt, Huaizhu Wu, Ron C. Hoogeveen, Nani Rebeiro, Joe L. Raya, Alice Lin, Kimberly Mackey, Massa Chen, and Mohamad Ghosn for their encouraging support, to Tracy Lynn Duncan for help with the manuscript, to Ellen M. Carroll (U.T.) and Rudy Guerra (R.U.) for reviewing its latter version, and to the *J.T.B.* reviewers and editors for enriching it. *Partly supported by NIH grant T32 HL 07812-11.

Appendix A

Codons, their properties, and exchangeability:

The codons for the essential (*) amino acids are

L (*)=*CUN, UUG, UUA*

V (*)=*GUN*²

T (*)=*ACN*

I (*)=*AUA, AUC, AUU*

F (*)=*UUU, UUC*

K (*)=*AAA, AAG*

M (*)=*AUG*

W (*)=*UGG*

Six of the essential amino acids are hydrophobic (M, I, V, L, F, and W), one is (T) and one is basic (K).

Codons ordered according to their number per amino acid 6 codons per amino acid (3)³

L (*)=*CUN, UUG, UUA*

R=*CGN, AGG, AGA*

S=*UCN, ACG, AGU*

4 codons per amino acid (5)

V (*)=*GUN*

T (*)=*ACN*

G=*GGN*

A=*GCN*

P=*CCN*

3 codons per amino acid or function (2)

I (*)=*AUA, AUC, AUU*

Stop=*UAA, UAG, UGA*

2 codons per amino acid (9)

F (*)=*UUU, UUC*

K (*)=*AAA, AAG*

E=*GAG, GAA*

D=*GAC, GAU*

Q=*CAG, CAA*

N=*AAC, AAU*

H=*CAU, CAC*

C=*UGU, UGC*

²An *N* means that any of the four nucleotides can be added here, which is the case for G, A, V and for R, P, L have a positional balance for quadrants 4 and 2, respectively; while also T and S have a positional balance for quadrants 3 and 1, respectively (a similar balancing effect is seen in the paired and in the odd-numbered groups of codons).

³Within the parenthesis is the number of amino acids (or function) within the category.

$$\mathbf{Y} = \text{UAC}, \text{UAU}$$

1 codon per amino acid (2)

$$\mathbf{M}^{(*)} = \text{AUG}$$

$$\mathbf{W}^{(*)} = \text{UGG}$$

Interchange of codons according to the rules of variation Interchange between the hydrophobic amino acids (the first two are the smaller amino acids)

$$\begin{array}{c} \mathbf{G} = \text{GGN} \leftrightarrow \mathbf{A} = \text{GCN} \leftrightarrow \mathbf{I}^{(*)} = \text{AUA, AUC, AUU} \leftrightarrow \mathbf{L}^{(*)} = \text{CUN, UUG, UUA} \leftrightarrow \mathbf{M}^{(*)} = \text{AUG} \\ \updownarrow \qquad \qquad \qquad \updownarrow \\ \mathbf{F} = \text{UUY} \leftrightarrow \mathbf{W}^{(*)} = \text{UGG} \quad \mathbf{V}^{(*)} = \text{GUN} \end{array}$$

Interchange between aromatic amino acids (the first two are hydrophobic)

$$\begin{array}{c} \mathbf{F} = \text{UUY} \leftrightarrow \mathbf{W}^{(*)} = \text{UGG} \\ \updownarrow \\ \mathbf{Y} = \text{UAC}, \text{UAU} \end{array}$$

Interchange between phosphorylatable amino acids

$$\mathbf{S} = \text{UCN, ACG, AGU} \leftrightarrow \mathbf{T}^{(*)} = \text{ACN}$$

The *square* of interchange between acidic amino acids and their amides

$$\begin{array}{ccc} \mathbf{E} = \text{GAG, GAA} & \leftrightarrow & \mathbf{D} = \text{GAC, GAU} \\ \updownarrow & & \updownarrow \\ \mathbf{Q} = \text{CAG, CAA} & \leftrightarrow & \mathbf{N} = \text{AAC, AAU} \end{array}$$

Interchange between basic amino acids

$$\mathbf{R} = \text{CGN, AGG, AGA} \leftrightarrow \mathbf{K}^{(*)} = \text{AAA, AAG}$$

Amino acids that barely interchange

$$\mathbf{H} = \text{CAU, CAC}$$

$$\mathbf{C} = \text{UGU, UGC}$$

$$\mathbf{P} = \text{CCN}$$

References

- Abe M, Kuroda Y, Hirose M, Kato M, Murakami M, Watanabe Y, Nakano M, Handa T. Inhibition of autophosphorylation of epidermal growth factor receptor by a small peptide not employing an ATP-competitive mechanism. *Biopolymers*. 2008; 89:40–51. Cited By in Scopus (3). [PubMed: 17849478]
- Alff-Steinberger C. The genetic code and error transmission. *Proc. Natl. Acad. Sci. USA*. 1969; 64:584–591. Cited By in Scopus (68). [PubMed: 5261915]
- Ambrogelly A, Palioura S, Soll D. Natural expansion of the genetic code. *Nat. Chem. Biol*. 2007; 3:29–35. Cited By in Scopus (36). [PubMed: 17173027]
- Armstrong M, Haldane F, Taylor RW, Humphriss D, Berrish T, Stewart MW, Turnbull DM, Alberti KG, Walker M. Human insulin receptor substrate-1: variant sequences in familial non-insulin-

- dependent diabetes mellitus. *Diabet. Med.* 1996; 13:133–138. Cited By in Scopus (13). [PubMed: 8641117]
- Arques DG, Michel CJ. A code in the protein coding genes. *BioSystems.* 1997; 44:107–134. Cited By in Scopus (12). [PubMed: 9429747]
- Arques DG, Lacan J, Michel CJ. Identification of protein coding genes in genomes with statistical functions based on the circular code. *BioSystems.* 2002; 66:73–92. Cited By in Scopus (2). [PubMed: 12204444]
- Arking DE, Reinier K, Post W, Jui J, Hilton G, O'Connor A, Prineas RJ, Boerwinkle E, Psaty BM, Tomaselli GF, Rea T, Sotoodehnia N, Siscovick DS, Burke GL, Marban E, Spooner PM, Chakravarti A, Chugh SS. Genome-wide association study identifies GPC5 as a novel genetic locus protective against sudden cardiac arrest. *PLoS One.* 2010; 5:e9879, 7. Cited By in Scopus (7). [PubMed: 20360844]
- Asensio JL, Siebert HC, von Der Lieth CW, Laynez J, Bruix M, Soedjanaamadja UM, Beintema JJ, Canada FJ, Gabius HJ, Jimenez-Barbero J. NMR investigations of protein–carbohydrate interactions: studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N''-triacylchitotriose. *Proteins.* 2000; 40:218–236. Cited By in Scopus (37). [PubMed: 10842338]
- Bag SK, Paul S, Ghosh S, Dutta C. Reverse polarization in amino acid and nucleotide substitution patterns between human-mouse orthologs of two compositional extrema. *DNA Res.* 2007; 14:141–154. Cited By in Scopus (3). [PubMed: 17895298]
- Barbieri M. Life is Semiosis: the biosemiotic view of nature. *Cosmos Hist. J. Nat. Soc. Philos.* 2008; 4:29–51. URL: <http://www.cosmosandhistory.org/index.php/journal/article/viewFile/99/197>.
- Basak S, Roy S, Ghosh TC. On the origin of synonymous codon usage divergence between thermophilic and mesophilic prokaryotes. *FEBS Lett.* 2007; 581:5825–5830. Cited By in Scopus (4). [PubMed: 18047839]
- Bashford D, Chothia C, Lesk AM. Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.* 1987; 196:199–216. Cited By in Scopus (224). [PubMed: 3656444]
- Bateson, W. *Mendel's Principles of Heredity: A Defence.* Cambridge University Press; 1902. p. 212 URL: <http://www.esp.org/books/bateson/mendel/facsimile>
- Belsham DD, Pereira F, Greenberg CR, Liao S, Wrogemann K. Leu-676-Pro mutation of the androgen receptor causes complete androgen insensitivity syndrome in a large Hutterite kindred. *Hum. Mutat.* 1995; 5:28–33. Cited By in Scopus (7). [PubMed: 7537149]
- Bellapadrona G, Chiaraluce R, Consalvi V, Ilari A, Stefanini S, Chiancone E. The mutations Lys 114→Gln and Asp 126→Asn disrupt an intersubunit salt bridge and convert *Listeria innocua* Dps into its natural mutant *Listeria monocytogenes* Dps. Effects on protein stability at low pH. *Proteins.* 2007; 66:975–983. Cited By in Scopus (9). [PubMed: 17186524]
- Berleant D, White M, Pierce E, Tudoreanu E, Boeszoermyenyi A, Shtridelman Y, Macosko JC. The genetic code—more than just a table. *Cell Biochem. Biophys.* 2009; 55:107–116. Cited By in Scopus (1). [PubMed: 19639425]
- Beyermann M, Fechner K, Furkert J, Krause E, Bienert M. A single-point slight alteration set as a tool for structure-activity relationship studies of ovine corticotropin releasing factor. *J. Med. Chem.* 1996; 39:3324–3330. Cited By in Scopus (44). [PubMed: 8765516]
- Biro JC, Benyo B, Sansom C, Szlavetz A, Fordos G, Micsik T, Benyo Z. A common periodic table of codons and amino acids. *Biochem. Biophys. Res. Commun.* 2003; 306:408–415. Cited By in Scopus (24). [PubMed: 12804578]
- Blikstad C, Shokeer A, Kurtovic S, Mannervik B. Emergence of a novel highly specific and catalytically efficient enzyme from a naturally promiscuous glutathione transferase. *Biochim. Biophys. Acta.* 2008; 1780:1458–1463. Cited By in Scopus (5). [PubMed: 18706975]
- Bosnacki, D.; Ten Eikelder, H.; Hibers, P. Genetic code as a gray code revisited. 2003 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences, METMBS; Las Vegas. 03, June 23–26; NV; 2003. p. 447–456.
- Bresch, C.; Hausmann, R. Der genetische code. In: Bresch, C.; Hausmann, R., editors. *Klassische und Molekulare Genetik.* Springer; Berlin: 1972. p. 243–278.

- Bunn HF, Noguchi CT, Hofrichter J, Schechter GP, Schechter AN, Eaton WA. Molecular and cellular pathogenesis of hemoglobin SC disease. *Proc. Natl. Acad. Sci. USA*. 1982; 79:7527–7531. Cited By in Scopus (22). [PubMed: 6961429]
- Castro-Chavez F, Yechoor VK, Saha PK, Martinez-Botas J, Wooten EC, Sharma S, O'Connell P, Taegtmeier H, Chan L. Coordinated upregulation of oxidative pathways and downregulation of lipid biosynthesis underlie obesity resistance in perilipin knockout mice: a microarray gene expression profile. *Diabetes*. 2003; 52:2666–2674. [Cited By in Scopus (27)]. URL: <http://diabetes.diabetesjournals.org/content/52/11/2666>. [PubMed: 14578284]
- Castro-Chavez F. Microarrays, antiobesity and the liver. *Ann. Hepatol*. 2004; 3:137–145. [Cited By in Scopus (3)]. URL: <http://www.medigraphic.com/pdfs/hepato/ah-2004/ah044c.pdf>. [PubMed: 15657555]
- Castro-Chavez F. Palindromati. *PCID*. 2005; 4.2:11. URL: http://www.iscid.org/pcid/2005/4/2/chavez_palindromati.php, <http://www.reocities.com/plin9k/t2.htm>.
- Castro-Chavez, F.; Vickers, KC.; Ghosn, M.; Reardon, M.; Lumsden, A.; Morrisett, JD. Celebrating BioScience Research Collaborative Grand Opening. Rice University; Houston: 2010. Osteogenic gene expression of vascular smooth muscle cells in atherosclerosis: Shn3's role. Poster Abstract Ten, Poster Session; p. 10April 1, 2010URL: <http://fdocc.ucoz.com/fdocc-rice.doc>
- Chen L, Li C, Qiao W, Xu X, Deng C. A Ser(365)→Cys mutation of fibroblast growth factor receptor 3 in mouse downregulates Ihh/PTHrP signals and causes severe achondroplasia. *Hum. Mol. Genet*. 2001; 10:457–465. Cited By in Scopus (61). [PubMed: 11181569]
- Chen XH, Tsien RW. Aspartate substitutions establish the concerted action of P-region glutamates in repeats I and III in forming the protonation site of L-type Ca^{2+} channels. *J. Biol. Chem*. 1997; 272:30002–30008. Cited By in Scopus (26). [PubMed: 9374474]
- Chen YP, Djaffar I, Pidard D, Steiner B, Cieutat AM, Caen JP, Rosa JP. Ser-752→Pro mutation in the cytoplasmic domain of integrin beta 3 subunit and defective activation of platelet integrin alpha IIb beta 3 (glycoprotein IIb-IIIa) in a variant of Glanzmann thrombasthenia. *Proc. Natl. Acad. Sci. USA*. 1992; 89:10169–10173. Cited By in Scopus (162). [PubMed: 1438206]
- Chiu DTY, Zuo L, Chao L, Chen E, Louie E, Lubin B, Liu TZ, Du CS. Molecular characterization of glucose-6-phosphate dehydrogenase (G6PD) deficiency in patients of Chinese descent and identification of new base substitutions in the human G6PD gene. *Blood*. 1993; 81:2150–2154. Cited By in Scopus (35). [PubMed: 8471773]
- Church GM. The personal genome project. *Mol. Syst. Biol*. 2005; 1:0030, 3. [PubMed: 16729065]
- Crick F. The origin of the genetic code. *J. Mol. Biol*. 1968; 38:367–379. Cited By in Scopus (491). [PubMed: 4887876]
- Crowder T, Li CK. Studying genetic code by a matrix approach. *Bull. Math. Biol*. 2010; 72:953–972. Cited By in Scopus (1). [PubMed: 20052557]
- Dalla-Venezia N, Alloisio N, Forissier A, Denoroy L, Aymerich M, Vives-Corrons JL, Besalduch J, Besson I, Delaunay J. Elliptopoikilocytosis associated with the alpha 469 His→Pro mutation in spectrin Barcelona (alpha I/50-46b). *Blood*. 1993; 82:1661–1665. [PubMed: 8364215]
- Davydov OV. Amino acid contribution to the genetic code structure: end-atom chemical rules of doublet composition. *J. Theor. Biol*. 1998; 193:679–690. (Article No. jt980733). [Cited By in Scopus (7)]. [PubMed: 9745760]
- Davis BK. On mapping the genetic code. *J. Theor. Biol*. 2009; 259:860–862. 10.1016/j.jtbi.2009.05.009. [Cited By in Scopus (2)]. [PubMed: 19463833]
- Dear A, Dempfle CE, Brennan SO, Kirschstein W, George PM. Fibrinogen Mannheim II: a novel gamma307 His→Tyr substitution in the gammaD domain causes hypofibrinogenemia. *J. Thromb. Haemost*. 2004; 2:2194–2199. Cited By in Scopus (10). [PubMed: 15613026]
- Demongeot J, Elena A, Weil G. Potential automata. Application to the genetic code III. *C.R. Biol*. 2006; 329:953–962. Cited By in Scopus (9). [PubMed: 17126799]
- Di Giulio M. Some aspects of the organization and evolution of the genetic code. *J. Mol. Evol*. 1989a; 29:191–201. Cited By in Scopus (33). [PubMed: 2506355]
- Di Giulio M. The extension reached by the minimization of the polarity distances during the evolution of the genetic code. *J. Mol. Evol*. 1989b; 29:288–293. Cited By in Scopus (42). [PubMed: 2514270]

- Di Giulio M. A comparison of proteins from *Pyrococcus furiosus* and *Pyrococcus abyssi*: barophily in the physicochemical properties of amino acids and in the genetic code. *Gene*. 2005; 346:1–6. Cited By in Scopus (18). [PubMed: 15716096]
- Dym O, Mevarech M, Sussman JL. Structural features that stabilize halophilic malate dehydrogenase from an archaeobacterium. *Science*. 1995; 267:1344–1346. Cited By in Scopus (141). [PubMed: 17812611]
- Elahi A, Bendaly J, Zheng Z, Muscat JE, Richie JP Jr, Schantz SP, Lazarus P. Detection of UGT1A10 polymorphisms and their association with orolaryngeal carcinoma risk. *Cancer*. 2003; 98:872–880. Cited By in Scopus (35). [PubMed: 12910533]
- Elstner M, Andreoli C, Ahting U, Tetko I, Klopstock T, Meitinger T, Prokisch H. MitoP2: an integrative tool for the analysis of the mitochondrial proteome. *Mol. Biotechnol.* 2008; 40:306–315. Cited By in Scopus (21). [PubMed: 18780189]
- Epstein CJ. Role of the amino-acid 'code' and of selection for conformation in the evolution of proteins. *Nature*. 1966; 210:25–28. Cited By in Scopus (48). [PubMed: 5956344]
- Epstein CJ. Non-randomness of amino-acid changes in the evolution of homologous proteins. *Nature*. 1967; 215:355–359. Cited By in Scopus (24). [PubMed: 4964553]
- Fan Y, Shi L, Brown LS. Structural basis of diversification of fungal retinal proteins probed by site-directed mutagenesis of *Leptosphaeria* rhodopsin. *FEBS Lett.* 2007; 581:2557–2561. Cited By in Scopus (4). [PubMed: 17499719]
- Feller G, Zekhnini Z, Lamotte-Brasseur J, Gerday C. Enzymes from cold-adapted microorganisms. The class C beta-lactamase from the antarctic psychrophile *Psychrobacter immobilis* A5. *Eur. J. Biochem.* 1997; 244:186–191. Cited By in Scopus (63). [PubMed: 9063463]
- Filatov, F. A molecular mass gradient is the key parameter of the genetic code organization. 2009. p. 9URL: <http://arxiv.org/pdf/0907.3537>
- Freeland SJ, Knight RD, Landweber LF, Hurst LD. Early fixation of an optimal genetic code. *Mol. Biol. Evol.* 2000; 17:511–518. Cited By in Scopus (99). [PubMed: 10742043]
- Gamow G. Possible relation between deoxyribonucleic acid and protein structures. *Nature*. 1954; 173:318. Cited By in Scopus (70).
- Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M. What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 2007; 17:669–681. Cited By in Scopus (126). [PubMed: 17567988]
- Goldberg AL, Wittes RE. Genetic code: Aspects of organization. *science*. 1966; 153:420–424. Cited By in Scopus (46). [PubMed: 5328568]
- Gomez-Casado E, Montoya F, Martinez-Laso J, Diaz-Campos N, Bedoya CI, Varela P, Arnaiz-Villena A. A novel HLA-B35 allele (B*3510) found in isolated Jaidukama Colombian American Indians. *Hum. Immunol.* 1996; 47:62.
- Hamosh A, King TM, Rosenstein BJ, Corey M, Levison H, Durie P, Tsui LC, McIntosh I, Keston M, Brock DJ, Macek M, Zemkova D, Krasnicanova H, Vavrova V, Macek M, Golder N, Schwarz MJ, Super M, Watson EK, Williams C, Bush A, O'Mahoney SM, Humphries P, DeArce MA, Reis A, Burger J, Stuhmann M, Schmidtke J, Wulbrand U, Dork T, Tummeler B, Cutting GR. Cystic fibrosis patients bearing both the common missense mutation Gly-Asp at codon 551 and the delta F508 mutation are clinically indistinguishable from delta F508 homozygotes, except for decreased risk of meconium ileus. *Am. J. Hum. Genet.* 1992; 51:245–250. Cited By in Scopus (33). [PubMed: 1379413]
- Hayes B. The invention of the genetic code [What mathematicians and physicists thought the code ought to be, before biologists deciphered the real one.]. *Am. Sci.* 1998; 86:9–14. Update: Chapter 4: Inventing the genetic code. In: *Group Theory in the Bedroom, and other Mathematical Diversions*. 2008, Hill and Wang, New York, 288 p. URL: <http://bit-player.org/bph-publications/AmSci-1998-01-Hayes-genetic-code.pdf>.
- Hayes B. Ode to the code [A closer look at the question of whether the code is optimal in some sense]. *Am. Sci.* 2004; 92:494–499. URL: <http://bit-player.org/bph-publications/AmSci-2004-11-Hayes-genetic-code.pdf>.
- Hayes P, Meadows HJ, Gunthorpe MJ, Harries MH, Duckworth DM, Cairns W, Harrison DC, Clarke CE, Ellington K, Ieiri I, Tainaka H, Morita T, Hadama A, Mamiya K, Hayashibara M, Ninomiya

- H, Ohmori S, Kitada M, Tashiro N, Prinjha RK, Barton AJ, Medhurst AD, Smith GD, Topp S, Murdock P, Sanger GJ, Terrett J, Jenkins O, Benham CD, Randall AD, Gloger IS, Davis JB. Cloning and functional expression of a human orthologue of rat vanilloid receptor-1. *Pain*. 2000; 88:205–215. Cited By in Scopus (170). [PubMed: 11050376]
- Hoffmeyer S, Burk O, von Richter O, Arnold HP, Brockmoller J, John A, Cascorbi I, Gerloff T, Roots I, Eichelbaum M, Brinkmann U. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc. Natl. Acad. Sci. USA*. 2000; 97:3473–3478. Cited By in Scopus (1334). [PubMed: 10716719]
- Humphries P, Farrar GJ, Kenna P, McWilliam P. Retinitis pigmentosa: genetic mapping in X-linked and autosomal forms of the disease. *Clin. Genet*. 1990; 38:1–13. Cited By in Scopus (20). [PubMed: 2201466]
- Huang DB, Brown EL, DuPont HL, Cerf J, Carlin L, Flores J, Belkind-Gerson J, Nataro JP, Okhuysen PC. Seroprevalence of the enteroaggregative *Escherichia coli* virulence factor dispersin among USA travellers to Cuernavaca, Mexico: a pilot study. *J. Med. Microbiol*. 2008; 57:476–479. Cited By in Scopus (3). [PubMed: 18349368]
- Ieiri I, Tainaka H, Morita T, Hadama A, Mamiya K, Hayashibara M, Ninomiya H, Ohmori S, Kitada M, Tashiro N, Higuchi S, Otsubo K. Catalytic activity of three variants (Ile, Leu, and Thr) at amino acid residue 359 in human CYP2C9 gene and simultaneous detection using single-strand conformation polymorphism analysis. *Ther. Drug Monit*. 2000; 22:237–244. Cited By in Scopus (65). [PubMed: 10850388]
- Jann OC, Werling D, Chang JS, Haig D, Glass EJ. Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. *BMC Evol. Biol*. 2008; 8:288. Cited By in Scopus (7). [PubMed: 18937834]
- Jestin JL, Kempf A. Optimization models and the structure of the genetic code. *J. Mol. Evol*. 2009; 69:452–457. Cited By in Scopus (4). [PubMed: 19841850]
- Jimenez-Montano MA, de la Mora-Basanez CR, Poschel T. The hypercube structure of the genetic code explains conservative and non-conservative amino acid substitutions *in vivo* and *in vitro*. *BioSystems*. 1996; 39:117–125. Cited By in Scopus (35). [PubMed: 8866048]
- Jimenez-Montano MA. The fourfold way of the genetic code. *BioSystems*. 2009; 98:105–114. Cited By in Scopus (2). [PubMed: 19643160]
- Jukes TH, Osawa S. The genetic code in mitochondria and chloroplasts. *Experientia*. 1990; 46:1117–1126. Cited By in Scopus (28). [PubMed: 2253709]
- Junca J, Villegas A, Ropero P, Gonzalez FA, Motos A, Valverde F. Characterization of a new hemoglobin variant: Hb Badalona (beta31[B13]Leu→Val). *Ann. Hematol*. 2002; 81:179–181. Cited By in Scopus (2). [PubMed: 11976817]
- Kane EM. Splitting the gene: DNA patents and the genetic code. *Tenn. Law Rev*. 2004; 71:707–767.
- Karasev VA, Stefanov VE. Topological nature of the genetic code. *J. Theor. Biol*. 2001; 209:303–317. doi:10.1006/jtbi.2001.2265 [Cited By in Scopus (13)]. [PubMed: 11312591]
- Karp, G. *Cell and Molecular Biology*. 5th ed. Wiley; New York: 2008. p. 73–74.
- Khorana HG. Polynucleotide synthesis and the genetic code. *Fed. Proc*. 1965; 24:1473–1487. Cited By in Scopus (6). [PubMed: 5322508]
- Khrennikova, A.Yu.; Kozyrev, SV. Genetic code on the diadic plane. *Physica A*. 2007; 381:265–272.
- Koene HR, Kleijer M, Algra J, Roos D, von dem Borne AE, de Haas M. Fc gammaRIIIa-158V/F polymorphism influences the binding of IgG by natural killer cell Fc gammaRIIIa, independently of the Fc gammaRIIIa-48L/R/H phenotype. *Blood*. 1997; 90:1109–1114. Cited By in Scopus (251). [PubMed: 9242542]
- Koolman, J.; Roehm, KH. *Thieme Color Atlas of Biochemistry*. 2nd ed. Thieme; Stuttgart: 2005. p. 249
- Lagerkvist U. “Two out of three”: an alternative method for codon reading. *Proc. Natl. Acad. Sci. USA*. 1978; 75:1759–1762. Cited By in Scopus (42). [PubMed: 273907]
- Lang R, Braun M, Sounni NE, Noel A, Franken F, Foidart JM, Bode W, Maskos K. Crystal structure of the catalytic domain of MMP-16/MT3-MMP: characterization of MT-MMP specific features. *J. Mol. Biol*. 2004; 336:213–225. Cited By in Scopus (22). [PubMed: 14741217]

- Larsen TB, Lassen JF, Brandslund I, Byriel L, Petersen GB, Nørgaard-Pedersen B. The Arg506Gln mutation (FV Leiden) among a cohort of 4188 unselected Danish newborns. *Thromb. Res.* 1998; 89:211–215. Cited By in Scopus (31). [PubMed: 9645914]
- Lecomte MC, Nicolas G, Dhermy D, Pinder JC, Gratzer WB. Properties of normal and mutant polypeptide fragments from the dimer self-association sites of human red cell spectrin. *Eur. Biophys. J.* 1999; 28:208–215. Cited By in Scopus (18). [PubMed: 10192935]
- Legato MJ. Father of the genetic code: thoughts on the life and death of Marshall W. Nirenberg. *Gend. Med.* 2010; 7:1–2. Cited By in Scopus (1). [PubMed: 20189148]
- Li BL, Li XL, Duan ZJ, Lee O, Lin S, Ma ZM, Chang CC, Yang XY, Park JP, Mohandas TK, Noll W, Chan L, Chang TY. Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes. *J. Biol. Chem.* 1999; 274:11060–11071. Cited By in Scopus (72). [PubMed: 10196189]
- Lorenzo F, Miraglia del Giudice E, Alloisio N, Morle L, Forissier A, Perrotta S, Sciaratta G, Iolascon A, Delaunay J. Severe poikilocytosis associated with a de novo alpha 28 Arg→Cys mutation in spectrin. *Br. J. Haematol.* 1993; 83:152–157. Cited By in Scopus (12). [PubMed: 8435324]
- Lu RC, Elzinga M. Partial amino acid sequence of brain actin and its homology with muscle actin. *Biochemistry.* 1977; 16:5801–5806. Cited By in Scopus (5). [PubMed: 588555]
- Linsel-Nitschke, P.; Goetz, A.; Lieb, W.; Doering, A.; Mayer, B.; Schunkert, H.; Erdmann, J.; Loewel, H. The 1527 Ala/Gly polymorphism in the ATP-binding cassette transporter ABCA7 is associated with low HDL-Cholesterol and metabolic syndrome in women. Abstract: P832. Poster session 1: Genetics and atherosclerosis. World Congress of Cardiology; Barcelona. 2–6 September; 2006. p. 7 URL: <http://spo.escardio.org/AbstractDetails.aspx?id=46939&eevtid=15>
- Lucio AD, Mazzoni MSC. Toxins by first-principles: electronic structure mapping structural changes. *J. Mol. Struct.* 2008; 853:58–61. Cited By in Scopus (1).
- Luo GX, Horowitz PM. The folding and stability of rhodanese are influenced by the replacement of glutamic acid 17 in the NH2-terminal helix by proline but not by glutamine. *J. Biol. Chem.* 1993; 268:10246–10251. Cited By in Scopus (10). [PubMed: 8098037]
- Lutsenko S, LeShane ES, Shinde U. Biochemical basis of regulation of human copper-transporting ATPases. *Arch. Biochem. Biophys.* 2007; 463:134–148. Cited By in Scopus (33). [PubMed: 17562324]
- Maeshiro T, Kimura M. The role of robustness and changeability on the origin and evolution of genetic codes. *Proc. Natl. Acad. Sci. USA.* 1998; 95:5088–5093. Cited By in Scopus (21). [PubMed: 9560233]
- Margoliash, E.; Smith, EL. Structural and functional aspects of cytochrome c in relation to evolution. In: Bryson, V.; Vogel, HJ., editors. *Evolving Genes and Proteins*. Academic Press; New York: 1965. p. 221–242.
- Martin RG, Matthaei JH, Jones OW, Nirenberg MW. Ribonucleotide composition of the genetic code. *Biochem. Biophys. Res. Commun.* 1962; 6:410–414. [PubMed: 14470439]
- Mendel, G. Ver. Natur. Verein. Brünn. Abhandlungen: 1865. Versuche über pflanzen hybriden; p. 3–47. Bd. IV für das Jahr 1865 English version—URL: <http://www.mendelweb.org/Mendel.html>
- Miller, KR.; Levine, JS. *Biology*; Prentice Hall. 2007. p. 1146 URL: <http://www.millerandlevine.com/images/Circular-Code.gif>
- Moura GR, Paredes JA, Santos MA. Development of the genetic code: insights from a fungal codon reassignment. *FEBS Lett.* 2010; 584:334–341. Cited By in Scopus (2). [PubMed: 19941859]
- Mundo E, Zai G, Lee L, Parikh SV, Kennedy JL. The 5HT1D(beta) receptor gene in bipolar disorder: a family-based association study. *Neuropsychopharmacology.* 2001; 25:608–613. Cited By in Scopus (14). [PubMed: 11557174]
- Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from the international DNA sequence databases. *Nucl. Acids Res.* 2000; 28:292. [Cited By in Scopus (399)]. URL: <http://www.kazusa.or.jp/codon>. [PubMed: 10592250]
- Negadi, T. Symmetry and proportion in the genetic code, and genetic information from the basic units of life. In: *Symmetries in genetic information*. In: Petoukhov, SV., editor. *Symmetry: Cult. Sci.* Vol. 12. 2001. p. 371–393. URL: <http://www.megaupload.com/?d=JKRXOTOH>, <http://symmetry.hu> Also see: http://www.biochempress.com/cv03_i05.html

- Nikolajewa S, Friedel M, Beyer A, Wilhelm T. The new classification scheme of the genetic code, its early evolution, and tRNA usage. *J. Bioinform. Comput. Biol.* 2006; 4:609–620. Cited By in Scopus (4). [PubMed: 16819806]
- Nikolajewa, S. Dissertation. Friedrich Schiller University of Jena; Germany: 2006. Identification and analysis of patterns in DNA sequences, the genetic code and transcriptional gene regulation; p. 102URL: <http://www.dbthueringen.de/servlets/DocumentServlet?id=6622>
- Oppermann UC, Filling C, Berndt KD, Persson B, Benach J, Ladenstein R, Jornvall H. Active site directed mutagenesis of 3 beta/17 beta-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions. *Biochemistry.* 1997; 36:34–40. Cited By in Scopus (90). [PubMed: 8993315]
- Pelc SR. Correlation between coding-triplets and amino-acids. *Nature.* 1965; 207:597–599. Cited By in Scopus (27). [PubMed: 5883631]
- Perria CL, Rajamanickam V, Lapinski PE, Raghavan M. Catalytic site modifications of TAP1 and TAP2 and their functional consequences. *J. Biol. Chem.* 2006; 281:39839–39851. Cited By in Scopus (8). [PubMed: 17068338]
- Plaseska D, de Alarcon PA, McMillan S, Walbrecht M, Wilson JB, Huisman TH. Hb Iowa or alpha 2 beta 2(119)(GH2)Gly---Ala. *Hemoglobin.* 1990; 14:423–429. [PubMed: 2283296]
- Pohlmeyer R. The genetic code revisited. *J. Theor. Biol.* 2008; 253:623–624. doi:10.1016/j.jtbi.2008.04.028 [Cited By in Scopus (3)]. [PubMed: 18519140]
- Qiu Y, Zhu L. The rearranged genetic code and its implications in evolution and biochemistry. *BioSystems.* 2000; 56:139–144. Cited By in Scopus (3). [PubMed: 10880861]
- Rakocevic MM. A harmonic structure of the genetic code. *J. Theor. Biol.* 2004; 229:221–234. doi: 10.1016/j.jtbi.2004.03.017 [Cited By in Scopus (3)]. [PubMed: 15207477]
- Ridgway HJ, Brennan SO, Fellowes AP, George PM. Rapid detection of the fibrinogen alpha 16Arg→His mutation. *Clin. Chem.* 1997; 43:2184–2186. Cited By in Scopus (3). [PubMed: 9365407]
- Ruiz-Rojas JJ, Sargent DJ, Shulaev V, Dickerman AW, Pattison J, Holt SH, Ciordia A, Veilleux RE. SNP discovery and genetic mapping of T-DNA insertional mutants in *Fragaria vesca* L. *Theor. Appl. Genet.* 2010; 121:449–463. [PubMed: 20349033]
- Rumer YB. About systematization of the codons of the genetic code. *Dokl. Acad. Nauk. SSSR.* 1966; 167:1393–1394.
- Saito K, Arai K, Mori M. p53 overexpression and K-ras codon 12 mutations in submucosal invasive depressed-type colorectal cancer. *Oncol. Rep.* 2000; 7:741–744. Cited By in Scopus (2). [PubMed: 10854536]
- Sanchez R, Grau R. A genetic code Boolean structure. II. The genetic information system as a Boolean information system. *Bull. Math. Biol.* 2005; 67:1017–1029. Cited By in Scopus (4). [PubMed: 15998493]
- Satoh J, Kuroda Y. A polymorphic variation of serine to tyrosine at codon 18 in the ubiquitin C-terminal hydrolase-L1 gene is associated with a reduced risk of sporadic Parkinson's disease in a Japanese population. *J. Neurol. Sci.* 2001; 189:113–117. Cited By in Scopus (57). [PubMed: 11535241]
- Serjeant GR, Serjeant BE. A comparison of erythrocyte characteristics in sickle cell syndromes in Jamaica. *Br. J. Haematol.* 1972; 23:205–213. Cited By in Scopus (3). [PubMed: 5070127]
- ShCherbak, V. The arithmetical origin of the genetic code. In: Barbieri, M., editor. *The Codes of Life. The Rules of Macroevolution.* Springer: 2008. p. 153–188. shortened URL: <http://tinyurl.com/y9w5gfj>, <http://fdocc.blogspot.com/2010/04/shcherbak-and-code.html>
- Shemirani AH, Muszbek L. Rapid detection of the factor XIII Val34Leu (163 G→T) polymorphism by real-time PCR using fluorescence resonance energy transfer detection and melting curve analysis. *Clin. Chem. Lab. Med.* 2004; 42:877–879. Cited By in Scopus (8). [PubMed: 15387436]
- Sheppard DN, Rich DP, Ostedgaard LS, Gregory RJ, Smith AE, Welsh MJ. Mutations in CFTR associated with mild-disease-form Cl-channels with altered pore properties. *Nature.* 1993; 362:160–164. Cited By in Scopus (195). [PubMed: 7680769]
- Shimada MK, Matsumoto R, Hayakawa Y, Sanbonmatsu R, Gough C, Yamaguchi-Kabata Y, Yamasaki C, Imanishi T, Gojobori T. VarySysDB: a human genetic polymorphism database

- based on all H-InvDB transcripts. Nucl. Acids Res. 2009; 37:D810–D815. [Cited By in Scopus (4)]. URL: <http://hinvitational.jp/varygene/home.htm>. [PubMed: 18953038]
- Shin JA, Kim SH, Kim UK, Chae JJ, Choe SJ, Namkoong Y, Kim HS, Park YB, Lee CC. Identification of four novel mutations of the low-density lipoprotein receptor gene in Korean patients with familial hypercholesterolemia. Clin. Genet. 2000; 57:225–229. Cited By in Scopus (4). [PubMed: 10782930]
- Siemion IZ, Siemion PJ. The informational context of the third base in amino acid codons. BioSystems. 1994; 33:139–148. Cited By in Scopus (17). [PubMed: 7811959]
- Sluyter R, Shemon AN, Wiley JS. Glu⁴⁹⁶ to Ala polymorphism in the P2×7 receptor impairs ATP-induced IL-1 beta release from human monocytes. J. Immunol. 2004; 172:3399–3405. Cited By in Scopus (62). [PubMed: 15004138]
- Smith WA, Hales BJ, Jarnicki AG, Thomas WR. Allergens of wild house dust mites: environmental Der p 1 and Der p 2 sequence polymorphisms. J. Allergy Clin. Immunol. 2001; 107:985–992. Cited By in Scopus (35). [PubMed: 11398075]
- Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TP, Sonstegard TS, Thallman RM, Bennett GL. Genome-wide association study of growth in crossbred beef cattle. J. Anim. Sci. 2010; 88:837–848. Cited By in Scopus (8). [PubMed: 19966163]
- SoRelle R. 1,000 genomes project: looking for the differences. Solutions. Baylor College of Medicine. 2008; 4(3):16–17. URL: <http://www.bcm.edu/solutions/v4i3/1000genomes.html>.
- Stafforini DM, Satoh K, Atkinson DL, Tjoelker LW, Eberhardt C, Yoshida H, Imaizumi T, Takamatsu S, Zimmerman GA, McIntyre TM, Gray PW, Prescott SM. Platelet-activating factor acetylhydrolase deficiency. A missense mutation near the active site of an anti-inflammatory phospholipase. J. Clin. Invest. 1996; 97:2784–2791. Cited By in Scopus (170). [PubMed: 8675689]
- Steinberg SF. Structural basis of protein kinase C isoform function. Physiol. Rev. 2008; 88:1341–1378. Cited By in Scopus (76). [PubMed: 18923184]
- Strachan, T.; Read, AP. Human Molecular Genetics. 2nd ed.. Wiley; New York: 1999. p. 576
- Swanson R. A unifying concept for the amino acid code. Bull. Math. Biol. 1984a; 46:187–203. Cited By in Scopus (58). [PubMed: 6733309]
- Swanson R. A vector representation for amino acid sequences. Bull. Math. Biol. 1984b; 46:623–639. Cited By in Scopus (5).
- Tradler T, Stoller G, Rucknagel KP, Schierhorn A, Rahfeld JU, Fischer G. Comparative mutational analysis of peptidyl prolyl cis/trans isomerases: active sites of *Escherichia coli* trigger factor and human FKBP12. FEBS Lett. 1997; 407:184–190. Cited By in Scopus (42). [PubMed: 9166896]
- Ueda Y, Ohwada S, Abe Y, Shibata T, Iijima M, Yoshimitsu Y, Koshiba T, Nakata M, Ueda T, Kawabata S. Factor G utilizes a carbohydrate-binding cleft that is conserved between horseshoe crab and bacteria for the recognition of beta-1,3-D-glucans. J. Immunol. 2009; 183:3810–3818. Cited By in Scopus (3). [PubMed: 19710471]
- van Bers NE, van Oers K, Kerstens HH, Dibbitts BW, Crooijmans RP, Visser ME, Groenen MA. Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. Mol. Ecol. 2010; 19(Suppl. 1):89–99. Cited By in Scopus (14). [PubMed: 20331773]
- Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: new opportunities for research and social challenges. Genome Med. 2010; 2:3. [PubMed: 20193048]
- Vickers K, Castro-Chavez F, Morrisett J. Abs. P252. Lysophosphatidylcholine induces osteogenic gene expression and phenotype in vascular smooth muscle cells. Atheroscler. Suppl. 2009; 10(2):e559. See also: <http://www.ncbi.nlm.nih.gov/pubmed/20451909> and http://content.onlinejacc.org/cgi/content/citation/57/14_Suppl_S/E1548.
- Viguera AR, Vega C, Serrano L. Unspecific hydrophobic stabilization of folding transition states. Proc. Natl. Acad. Sci. USA. 2002; 99:5349–5354. Cited By in Scopus (53). [PubMed: 11959988]
- Vogtherr M, Jacobs DM, Parac TN, Maurer M, Pahl A, Saxena K, Ruterjans H, Griesinger C, Fiebig KM. NMR solution structure and dynamics of the peptidyl–prolyl cis–trans isomerase domain of the trigger factor from *Mycoplasma genitalium* compared to FK506-binding protein. J. Mol. Biol. 2002; 318:1097–1115. Cited By in Scopus (21). [PubMed: 12054805]

- Voie OA. Biological function and the genetic code are interdependent. *Chaos Solitons Fractals*. 2006; 28:1000–1004. Cited By in Scopus (1).
- Wang B, Wang J, Liu J, Ni F, Yan J, Zhou S, Mu Y, Cao Y, Ma X. Lack of an association between CYP1A1 gene Ile462Val polymorphism and polycystic ovary syndrome in Chinese. *Endocrine*. 2009; 36:16–19. (in Chinese). [Cited By in Scopus (1)]. [PubMed: 19507064]
- Wang JD, Herman C, Tipton KA, Gross CA, Weissman JS. Directed evolution of substrate-optimized GroEL/S chaperonins. *Cell*. 2002; 111:1027–1039. Cited By in Scopus (53). [PubMed: 12507429]
- Wang X. Molecular characterization of a novel defect occurring de novo associated with erythropoietic protoporphyria. *Biochim. Biophys. Acta*. 1996; 1316:149–152. Cited By in Scopus (8). [PubMed: 8781532]
- White HB III, Dhurjati P. Evolution of protein lipograms a bioinformatics problem. *Biochem. Mol. Biol. Educ*. 2006; 34:262–266. Cited By in Scopus (2). [PubMed: 21638688]
- Wittmann, HG. Studies on the genetic code in tobacco mosaic virus. In: Vogel, HJ.; Bryson, V.; Lampen, JO., editors. *Informational Macromolecules: A Symposium*. Academic Press; New York: 1963. p. 177–193.
- Woese CR. Order in the genetic code. *Proc. Natl. Acad. Sci. USA*. 1965; 54:71–75. Cited By in Scopus (40). [PubMed: 5216368]
- Woese CR, Dugre SA, Dugre DH, Kando M, Saxinger WC. On the fundamental nature and evolution of the genetic code. *Cold Spring Harb. Symp. Quant. Biol*. 1966; 31:723–736. Cited By in Scopus (79). [PubMed: 5237212]
- Worn A. An intrinsically stable antibody scFv fragment can tolerate the loss of both disulfide bonds and fold correctly. *FEBS Lett*. 1998; 427:357–361. Cited By in Scopus (77). [PubMed: 9637257]
- Yang CM. On the 28-gon symmetry inherent in the genetic code intertwined with aminoacyl-tRNA synthetases—the Lucas series. *Bull. Math. Biol*. 2004; 66:1241–1257. Cited By in Scopus (2). [PubMed: 15294424]
- Yanofsky C, Ito J, Horn V. Amino acid replacements and the genetic code. *Cold Spring Harb. Symp. Quant. Biol*. 1966; 31:151–162. Cited By in Scopus (9). [PubMed: 4866372]
- Yockey, HP. *Information Theory, Evolution, and the Origin of Life*. Cambridge University Press; New York: 2005. p. 98
- Yu J. A content-centric organization of the genetic code. *Genom. Proteom. Bioinf*. 2007; 5:1–6. Cited By in Scopus (5).
- Zhu W, Freeland S. The standard genetic code enhances adaptive evolution of proteins. *J. Theor. Biol*. 2006; 239:63–70. doi:10.1016/j.jtbi.2005.07.012 [Cited By in Scopus (6)]. [PubMed: 16325205]

Rule 1. Same position in the four quadrants for essential hydrophobic codons (U at the center)

Fig. 1.

Schematic representation of Rule 1. The seven hydrophobic amino acids are encased inside a rectangle and may be considered the engines of variation. The hydrophobic essential amino acids Val, Met, Ile, Phe, and Leu have codons with U (T for DNA) at its central position (encircled). There are 21 hydrophobic codons. Less frequent exchanges are indicated by intermittent rectangles (circular genetic code modified from Miller and Levine, 2007). The arrows represent rotational ability of the genetic code for its use as a reproducible clinical tool.

Rule 2. Same or contiguous position in two quadrants for synonymous or related codons

Fig. 2.

Schematic representation of Rule 2. Identical amino acids are arginine in quadrants two and three, lysine in quadrants one and two, and serine in quadrants one and three. Amino acids that are closely related are aspartic acid and asparagine in quadrants four and three, glutamic acid and glutamine in quadrants four and two, and glutamine and asparagine among themselves. Only some examples are illustrated. The nucleotide at the center is encircled (circular genetic code modified from Miller and Levine, 2007).

Rule 3. Same quadrant for equivalent codons **Fig. 3.**

Schematic representation of Rule 3. In the same quadrant there is similarity of codons for equivalent amino acids or of codons with similar function: acidic amino acids (Asp, Glu), hydrophobic (Ala, Val) and small amino acids (Gly, Ala) in quadrant four, basic amino acids (Arg, Lys) and highly phosphorylatable amino acids (“signal transducers”) in quadrant three (Ser, Thr), and aromatic amino acids (Phe, Trp, Tyr) and stop codons in quadrant one. Only some examples are illustrated (circular genetic code modified from Miller and Levine, 2007). Based on these three rules of variation, some amino acid exchanges are exemplified (Fig. 4).

Fig. 4.

The arrows indicate that the amino acids may be reciprocally interchangeable while retaining protein function: Ala for Val, Val for Ile or Leu, Leu, and Met, while the aromatic Phe, Trp, and Tyr seem to undergo some limited exchange. The two smallest amino acids seem to be interchangeable: Gly for Ala. In the polar charged amino acids the permissible exchanges seem to be Glu for Asp and Arg for Lys. Also, it is possible to have either a Ser or a Thr in the same residue position while retaining the protein function, and the same goes for Gln and Asn. There is also the exchangeability of Gln by Glu and Asn by Asp. In some exceptional instances Ala can be replaced by Phe and Phe by Ile; however, His inside the polar charged amino acids seems to be more limited in its exchange, as well as Cys connecting peptides and Pro for the protein's shape (original drawing done using *Word™*). A more restricted exchange and effect may be experienced depending on their location within the protein (according to *VarySysDB*, by Shimada et al., 2009).