DNA, Amino Acids and the Digital Root

Dr. Talal Ghannam

In this research paper I will demonstrate how numbers resides at the core of the DNA itself. This I do by explaining some of the most intriguing and puzzling aspects of the coding process of the DNA by the implementing one of the most simple mathematical operations; namely the Digital Root, which is will be tackled at the last part of this paper.

Simply, the digital root of a number is the operation of adding the individual digits of a number until it is reduced to a single digit.

So for example, for a number like 512, the digital root is $D(512) = 5+1+2 = 8$. (Here D stands for the operation of taking the digital root.) Another example is 137: $D(137) = D(1+3+7) = D(11) = 1+1 = 2$.

The term D-sum refers to applying the digital root operation for the sum of two numbers. So the D-sum of 512 and 137 is $D(137+512) = D(649) = D(19) = 1$. (In this world of the digital root, number 9 acts as number 0: $D(9+4) = D(13) = 1+3 = 4$.)

It is amazing how an extremely simple operation like this one can reveal so much about the universe as detailed in my book: The Mystery of Numbers Revealed through their Digital Root.

For this paper, I will apply this method to reveal some of the logic and symmetries that reside hidden deep within the DNA code.

Nucleotides: The Basic Units of Life.

DNA are made of a set of four chemical compounds called nucleotides; Alanine (G), Guanine (G), Cytosine (C), and Thymine (T). In the RNA, thymine is replaced by uracil (U). RNA is considered unstable when compared to the DNA. It is always the case that alanine combines with thymine, and guanine with cytosine; A-T, and G-C. These five nucleotides are made of four main elements only: Hydrogen, Carbon, Nitrogen, and Oxygen.

The following table list the five nucleotides with their chemical compositions.

Nucleotide	Chemical Composition	Total Number of Elements	*D* (Digital Root)
Alanine (A)	$C_5H_5N_5$	15	6
Thymine (T)	$C_5H_6N_2O_2$	15	6
Guanine (G)	$C_5H_5N_5O_1$	16	7
Cytosine (C)	$C_4H_5N_3O_1$	13	4
Uracil (U)	$C_4H_4N_2O_2$	12	3

A list of the five nucleotides along with their chemical composition.

From the picture shown below, we can see how all the nucleotides are based on a hexagonal-pentagonal design, in other words 5 and 6.

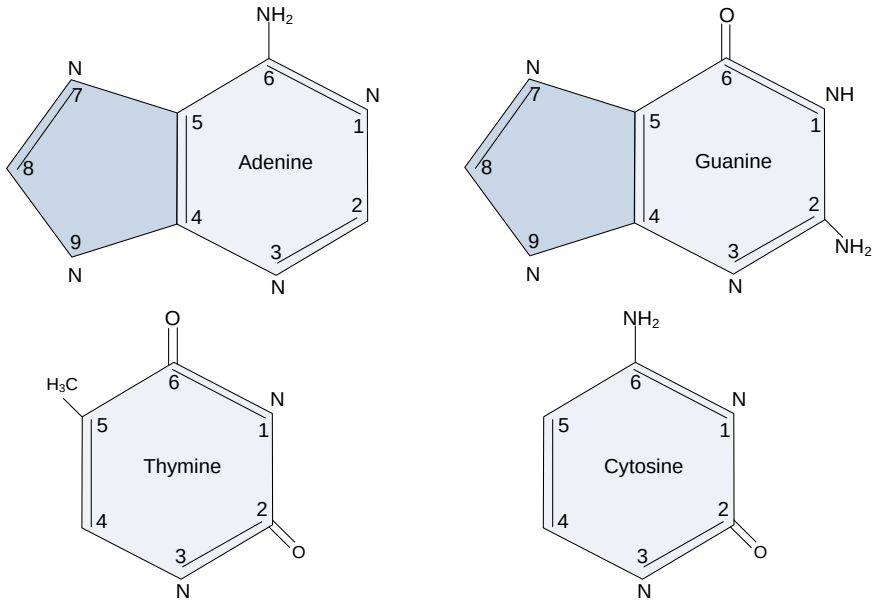


Figure 1: The four main nucleotides.

The overall helical structure of the DNA is supported by a shell or double rods made of sugar and phosphate, making the total number of basic building units involved in the DNA structure equals to 6, a number, along with 5, seem to define the whole structure of the DNA, as we to discover later on.
In the DNA, there are two strands; sense and anti-sense. Both strands can be used to code for amino acids etc. The RNA structure, on the other hand, is made of one strand that is usually not helical in shape.

Numbering Schemes for the Nucleotides

In order to implement the digital root in the analysis of the DNA, we need first to put forth a numbering mechanism that will numerically codify the elements of the system, the nucleotides in this case.
There are several methods one can think of to assign numbers to the five nucleotides. In this report I will adopt that method which is based on the electronic configuration of elements involved in the composition of the nucleotides.
Below I list two different numbering schemes derived from the above mentioned method.

1- Based on the number of electrons in each element.

Nucleotide	H	C	N	O	Sum of e	*D*	By-Sum	*D*-sum
A	5	30	35	0	70	7	10	1
T	6	30	14	16	66	3		
G	5	30	35	8	78	6	10	1
C	5	24	21	8	58	4		
U	4	24	14	16	58	4	11	2

Notice how in this scheme, numbers belong mainly to the [1, 4, 7] and [3, 6, 9] triplet groups: [7, 3], [6, 4] and 1 for their *D*-sum. (The Triplet groups are a unique and fundamental distribution of numbers into the groups [1, 4, 7], [2, 5, 8] and [3, 6, 9], and are explained in detailed in my book: The Mystery of Numbers Revealed through their Digital Root)

2- Based on the number of electrons of the outer shell only.
This method rests on geometrical as well as on atomic and chemical reasoning as the electrons of the outer shells are the ones responsible for chemical properties of the element. In addition, they form geometrical patterns which may very well be involved in how atoms interact with each other.

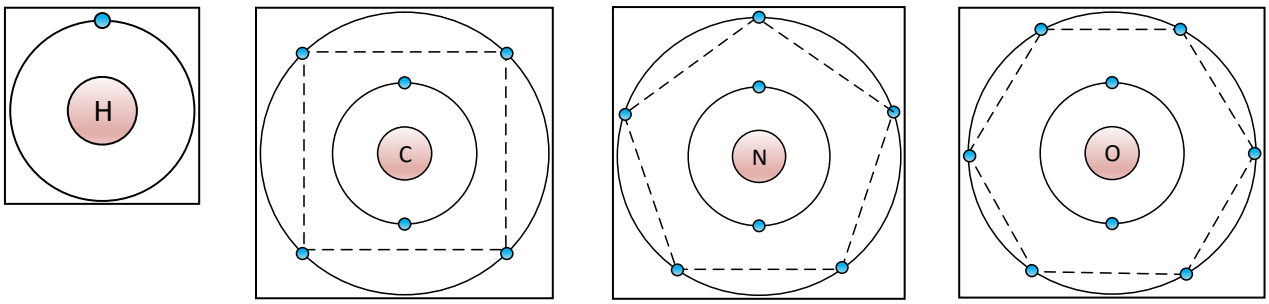


Figure 2. The main four elements making geometrical shapes of their outer shells, namely a point, a square, a pentagon, and a hexagon.

The numerical table corresponding to the above scheme is shown below.

Nucleotide	H	C	N	O	Sum of e	*D*	By-Sum	*D*
A	5	20	25	0	50	5	8	8
T	6	20	10	12	48	3		
G	5	20	25	6	56	2	8	8
C	5	16	15	6	42	6		
U	4	16	10	12	42	6	2	2

Interestingly, this scheme involves numbers [2, 3, 5, 6], and with the By-sum of the nucleotides T-A and C-G being equal 8 for both.
Notice also how in this scheme, A and T are odd [5, 3] and G and C are even [6, 2].

There are many other numbering schemes one can use, in addition to the above ones. However, I believe that the most logical and promising one is scheme two, which is the one I will adopt in my research.

About Numbers [5, 3, 6, 2] and their Relation to the DNA

Interestingly enough, most of these numbers, [5, 3, 6, 2], are embedded deeply within the design of the DNA.
First of all, the DNA is made of double strands, which implies polarity and duality, in other words number 2.
The structure of the nucleotides is made of pentagonal and hexagonal rings, which are numbers 6 and 5, as was shown above in figure 1. Moreover, the total number of basic elements involved in making up the DNA is 5: H, P, O, N, and C. (Interestingly, phosphorus has 15 electrons, with a $D = 6$, arranged in three shells of [2, 8, 5], which is very significant in its own right, as far as the basic triplet is concerned).
There are 5 main nucleotides involved in the coding process. However, sometimes a 6^{th} one is involved, called *inosine*, found only in the anti-codon of the tRNA, which is attached to the amino acid and also having a pentagonal-hexagonal structure like adenine and guanine. Hence again 5 and 6.
Numbers 5 and 3 are found in the structure of the DNA chains which have 5 free phosphate or hydroxyl at one end and 3 free phosphate or hydroxyl at the other end. In other words, the base at the 5' end of one strand is paired with the base at the 3' end of the other strand.

Digital-root speaking, numbers 5 and 6, by themselves, are able to generate the other two numbers, 2 and 3, as $5+6 = 11$ and $D(11) = 2$. And $5\times6 = 30$ and $D(30) = 3$. Moreover, $2\times3 = 6$ and $2+3 = 5$.

When it comes to geometry, the pentagon and the hexagon are very intrinsically connected. Both can be generated, and with an almost exact equal sides (1.7% difference), from the same circle as shown below to the left, or exactly from the Vesica Piscis symbol, to the right.

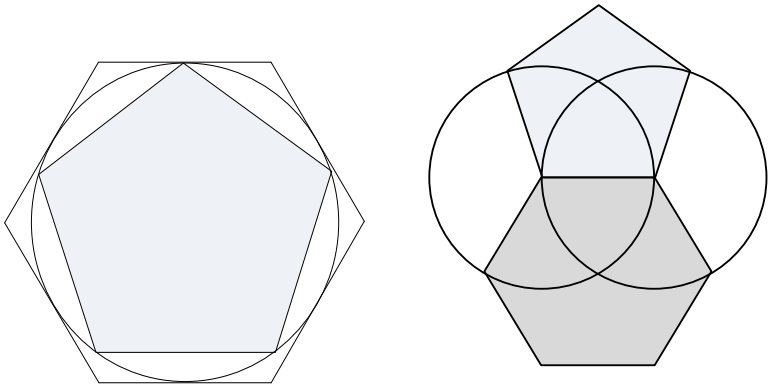


Figure 4: Left: Two equal sides pentagon and the hexagon define each other via the same middle circle. Right: The Vesica Pisces defines two equal sides pentagon and hexagon. (Notice that 6+5 =2 and number 2 is, in my own understanding, the numerical value of the circle.)

Hence, to get these specific numbers, numbers that are at the core of the DNA and life, emerging straightforwardly from the logical assignment of numbers to the atoms that made the nucleotides, is very thought provoking and profound indeed.

Some Interesting Facts About 5 and 6

1. Number 6 is the structuring number, can be found everywhere from molecules to crystals symmetries, to bee hives, to Saturn's norther pole etc.
2. Number 5 has always been consider the number of man and life. Almost all edible plants have 5-petal roses. It is the number that defines $\Phi = 0.5 + 0.5\times5^{0.5} = 1.618\ldots$ were Φ is the famous golden section found everywhere in nature from atoms, to the DNA itself, to the largest galaxies.
3. The pentagon can form a 3-Dimesnianl closed object (the dodecahedron), the hexagon cannot.
4. The hexagon can fill out the whole 2-Dimensional space, but cannot form a 3-Dimensional object.
5. Together, they connect the most important constant of nature together, Φ and π, through this wonderful equation: $6/5 \times \Phi^2 = \pi$.
6. Water molecules cluster themselves in pentagonal shapes, especially around living proteins. And when cooled to freezing temperatures they form hexagonal structures. Hence, water is the substance that can capture both essence of nature: the hexagon and the pentagon, 2-dimentional and 3 dimensional, etc.

The Amino Acids and their Codons

Codons are groups of three nucleotides bundled together. They are used to code for the amino acids, which in turn are used to build the proteins that make up the whole body.
This is done by the mRNA, which copies the required sequence from one strand of the DNA, which then is used to build the necessary amino acid.

The four different nucleotides, A, T,, C, and G, can form 64 different codons. These codons are then used to code for the 20 different amino acids plus extra 3, which are the stop codons that end the coding or copying process and hence inform the RNA to stop the process of creating the amino acid. So in total, we have 23, with a *D* of 5.

When each nucleotides in every codon is given a numerical value taken from scheme two, [A, T, G, C] = [5, 3, 2, 6], and then the digital root of the entire codon value is calculated (*D*-sum), the codons will occupy the following table with the numbers on the top representing the *D*-sum of the three nucleotides.

1	2	3	4	5	6	7	8	9
TAG	TTA	TTC	TAA	TCA	TCC	TGG	TTG	TTT
TGA	TCG	TCT	CAG	TAC	CTC	CAA	TGT	CCC
CGG	TAT	CTT	CGA	CTA	CCT	ACA	CCA	AGG
ATG	TGC	AAG	ATA	CCG	AAA	AAC	CAC	GAG
AGT	CTG	AGA	ACG	CAT	GGG	GTG	ACC	GGA
GTA	CGT	GAA	AAT	CGC		GGT	GTT	
GCG	ATT		AGC	ATC				
GAT	GTC		GCA	ACT				
GGC	GCT		GAC	GCC				

For example, the TAG codon has a digital root equal to 1 which comes from *D*(T+A+G) = *D*(3+5+2) = *D*(10) = 1. Hence, all the codons that falls under the same column share the same digital root.
Notice how no two codons belonging to *the same column*, in other words having the *same digital root*, code for the same amino acid, except for the case of Leucine. This I found to be a very interesting behavior, which will be explained at the end of this report.

The codons are distributed in an ostensible random fashion between the amino acids. Some amino acids have three codons, some four, some two etc. Now, I will call these codons belonging to the same amino

acid, the *codonic family* of that amino acid, so for example the codonic family of alanine, or the codonic family of proline, etc.

The different amino acids along with their corresponding codons are shown in the table below:

Amino Acid	Codonic Family: Sense Strand	The Digital Root for each Codon						Sum	*D*-Sum
Isoleucine	ATT, ATC, ATA	2	5	4				11	2
Leucine	TTA, TTG, CTT, CTC, CTA, CTG	2	8	3	6	5	2	26	8
Valine	GTT, GTC, GTA, GTG	8	2	1	7			18	9
Phenylalanine	TTT, TTC	9	3					12	3
Methionine	ATG	1						1	1
Cysteine	TGT, TGC	8	2					10	1
Alanine	GCT, GCC, GCA, GCG	2	5	4	1			12	3
Glycine	GGT, GGC, GGA, GGG	7	1	9	6			23	5
Proline	CCT, CCC, CCA, CCG	6	9	8	5			28	1
Threonine	ACT, ACC, ACA, ACG	5	8	7	4			24	6
Serine	TCT, TCC, TCA, TCG, AGT, AGC	3	6	5	2	1	4	21	3
Tyrosine	TAT, TAC	2	5					7	7
Tryptophan	TGG	7						7	7
Glutamine	CAA, CAG	7	4					11	2
Asparagine	AAT, AAC	4	7					11	2
Histidine	CAT, CAC	5	8					13	4
Glutamic Acid	GAA, GAG	3	9					12	3
Aspartic Acid	GAT, GAC	1	4					5	5
Lysine	AAA, AAG	6	3					9	9
Arginine	AGA, AGG, CGT, CGC, CGA, CGG	3	9	2	5	4	1	24	6
Stop Codons	TAA, TAG, TGA	4	1	1					6
								291	93

Amino Acid	Codonic Family: Anti-sense Strand	The Digital Root for each Codon						Sum	*D*-Sum
Isoleucine	TAA, TAG, TAT	4	1	2				7	7
Leucine	AAT, AAC, GAA, GAG, GAT, GAC	4	7	3	9	1	4	28	1

Valine	CAA, CAG, CAT, CAC	7	4	5	8			24	6
Phenylalanine	AAA, AAG	6	3					9	9
Methionine	TAC	5						5	5
Cysteine	ACA, ACG	7	4					11	2
Alanine	CGA, CGG, CGT, CGC	4	1	2	5			12	3
Glycine	CCA, CCG, CCT, CCC	8	5	6	9			28	1
Proline	GGA, GGG, GGT, GGC	9	6	7	1			23	5
Threonine	TGA, TGG, TGT, TGC	1	7	8	2			18	9
Serine	AGA, AGG, AGT, AGC, TCA, TCG	3	9	1	4	5	2	24	6
Tyrosine	ATA, ATG	4	1					5	5
Tryptophan	ACC	8						8	8
Glutamine	GTT, GTC	8	2					10	1
Asparagine	TTA, TTG	2	8					10	1
Histidine	GTA, GTG	1	7					8	8
Glutamic Acid	CTT, CTC	3	6					9	9
Aspartic Acid	CTA, CTG	5	2					7	7
Lysine	TTT, TTC	9	3					12	3
Arginine	GCA, GCG, GCT, GCC, TCT, TCC	4	1	2	5	3	6	21	3
Stop Codons	ATT, ATC, ACT	2	5	5				12	3
								291	102

There are some very interesting observations one can make from the above two tables:

1- There are no amino acids with *5 codons*.

2- The maximum number of codons coding for the same amino acid is 6.

3- The sense and anti-sense codons will add up to number 6.

4- The codons comes in groups of two, two, with the first two belonging to the members of one group of the triplet, followed by another two of another triplet and so on. So for Valine, we have [7, 4, 5, 8], [7, 4] belongs to the [1, 4, 7] group and [5, 8] belong to the [2, 5, 8] group.

In the below table I list the amino acids codons of the sense strand (1st) and their opposites in the anti-sense strand (2nd) strand, along with their numerical values.

Strand		DNA Codons	1	2	3	4	5	6	D-Sum	*D*-sum
1st	Isoleucine	ATT, ATC, ATA	2	5	4				2	9
2nd		TAA, TAG, TAT	4	1	2				7	

1st	Leucine	TTA, TTG, CTT, CTC, CTA, CTG	2	8	3	6	5	2	8	9
2nd		AAT, AAC, GAA, GAG, GAT, GAC	4	7	3	9	1	4	1	
1st	Valine	GTT, GTC, GTA, GTG	8	2	1	7			9	6
2nd		CAA, CAG, CAT, CAC	7	4	5	8			6	
1st	Phenylalanine	TTT, TTC	9	3					3	3
2nd		AAA, AAG	6	3					9	
1st	Methionine	ATG	1						1	6
2nd		TAC	5						5	
1st	Cysteine	TGT, TGC	8	2					1	3
2nd		ACA, ACG	7	4					2	
1st	Alanine	GCT, GCC, GCA, GCG	2	5	4	1			3	6
2nd		CGA, CGG, CGT, CGC	4	1	2	5			3	
1st	Glycine	GGT, GGC, GGA, GGG	7	1	9	6			5	6
2nd		CCA, CCG, CCT, CCC	8	5	6	9			1	
1st	Proline	CCT, CCC, CCA, CCG	6	9	8	5			1	6
2nd		GGA, GGG, GGT, GGC	9	6	7	1			5	
1st	Threonine	ACT, ACC, ACA, ACG	5	8	7	4			6	6
2nd		TGA, TGG, TGT, TGC	1	7	8	2			9	
1st	Serine	TCT, TCC, TCA, TCG, AGT, AGC	3	6	5	2	1	4	3	9
2nd		AGA, AGG, AGT, AGC, TCA, TCG	3	9	1	4	5	2	6	
1st	Tyrosine	TAT, TAC	2	5					7	3
2nd		ATA, ATG	4	1					5	
1st	Tryptophan	TGG	7						7	6
2nd		ACC	8						8	
1st	Glutamine	CAA, CAG	7	4					2	3
2nd		GTT, GTC	8	2					1	
1st	Asparagine	AAT, AAC	4	7					2	3
2nd		TTA, TTG	2	8					1	
1st	Histidine	CAT, CAC	5	8					4	3
2nd		GTA, GTG	1	7					8	
1st	Glutamic Acid	GAA, GAG	3	9					3	3
2nd		CTT, CTC	3	6					9	
1st	Aspartic Acid	GAT, GAC	1	4					5	3
2nd		CTA, CTG	5	2					7	
1st	Lysine	AAA, AAG	6	3					9	3
2nd		TTT, TTC	9	3					3	
1st	Arginine	CGT, CGC, CGA, CGG, AGA, AGG	3	9	2	5	4	1	6	9
2nd		GCA, GCG, GCT, GCC, TCT, TCC	3	6	4	1	2	5	3	

We can see how in every pair of strands, the [2, 5, 8] group pairs only with the [1, 4, 7] one, and the [3, 6, 9] group pairs with itself.

The last two columns of the above table may not mean much, as the codons generally do not form a one unified block. Nevertheless, it is still interesting to see the numerical patterns that emerges from treating them together.
For example, when adding both strands, the triple codons on each strand will always add up to 6. This is expected as we have already chosen our number scheme such that adding two nucleotides will result in 8, so three of them will give 24 and hence $D(24) = 6$. For the same reason, the total codons for each amino acid will all add up to 3, 6, and 9 depending on the number of codons in each amino acids.

The Codonic Family and their Order

Before we proceed in the numerical analysis of the amino acids, I need to say something first about the reasoning behind the above order of the codons.
It is generally believed that each codon belonging to the same amino acid will code for this same amino acid alone, and without any need for the other codons in the family, and hence treating the codonic family of a certain amino acid as a single unit does not make sense. Moreover, ordering them in a precise and consistent way is not necessary either.
Nevertheless, and as we are to discover later on, when considering the numerical behavior of these codons, the codonic family of each amino acid seem to act as one unified block, this is because when treated as such, they will reveal amazing hidden information about themselves, information that cannot be deduced at all if they were treated individually.

Moreover, the specific order of the nucleotides, being (XX)**T**, (XX)**C**, (XX)**A**, and (XX)**G,** has its merits also. As when looking at the amino acids that have two codons in their codonic family, we never see something like (XX)A, (XX)T. The third nucleotides of the pair codons Always comes either T and C or A and G. (This will also ensure that the couples always come from the same triplet group.) So for example, Phenylalanine has the following codonic family: TTT, TTC [9, 3]. Tyrosine is TAT and TAC [2, 5], etc. We don't see TTT and TAG [9, 1] for example or AGT and AGA [3, 1].

Now, if the distribution of the codons between the various amino acids is totally random, why do the codons follow the above pattern for each amino acids? Why don't I have something like AAG and CTG for phenylalanine for example? There must be a reason behind it. There is something about the first two

nucleotides making them very important to the *identity* coding of the amino acid. The third nucleotide seems to be not as important as the first two. This conclusion is also justifiable from a chemical point view as the third nucleotides are sometimes less bonded between the messenger RNA and he anti codons, than the first two are. This happens when the third codon in the RNA and the third anti-codon attached the amino acids do not match. (This behavior is called the wobble base pair hypothesis.)

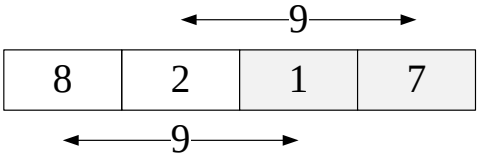
Numerical Patterns and Conditions

By observing the codonic family of each amino acid as a whole, some basic numerical patterns and laws can be deduced. In all of these patterns, we will be looking at what makes the system more ordered and with the least number of variables. Even though these laws may seem unimportant at first, when we remember that the codons are actually independent entities, nevertheless, later on when we try to understand the mysterious codonic distribution between the different amino acids, these laws will turn out to be important and relevant.

As was said earlier, and in order for us to observe these numerical laws, the codons in each codonic family should be ordered systematically such that the third nucleotide of every codonic family will adopt the following sequence: [T, C, A, G], partially or entirely. I will call this condition the *codons sequence condition*.
So, for valine, for example, the sequence will be GT**T**, GT**C**, GT**A**, GT**G**, and so on for the rest of the amino acids.
For those amino acids having six codons, the extra two codons will share the same sequence condition like any other amino acids that does not have four codons.
So for Serine, for example, the first four will share the nucleotides TC, while the remaining two will share AG. This is logical, for as long as we only have four nucleotides bases, we can only have four variations of codons while sharing the first two nucleotides at the same time: (XX)**T**, (XX)**C**, (XX)A, and (XX)**G**.

The first numerical consequence of the numbers assigned to the nucleotides is that the codons will come in couples belonging to the same triplet, like [3, 6], [5, 2], [1, 4] for Serine.

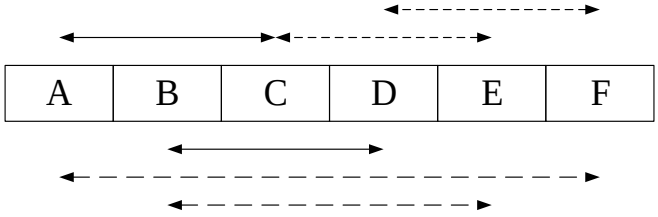
The codons sequence condition ensures that the first four codons are ordered such that the sum of the first and third codon equals to the sum of the second and fourth. So for example, for alanine in the sense strand, the numerical values of the codons are [2, 5, 4, 1] and we find that 2+4 = 5 +1 = 6.



Moreover, always the 3rd codon is 1 less than the 2nd and the 4th is 1 less than the 1st. So for alanine, the 2nd and third are 5 and 4, and the 1st and 4th are 2 and 1.

For the anti-sense strand, the rule is flipped around with the 3rd codon having one numerical value bigger than the 2nd, and so is for the 1st and 4th.

For the 6-codon-based amino acids, the following rules apply (the letters refer to different codons, not to be confused with nucleotides):



1- A + C = B + D
2- C + E = D + F
3- A + F = B + E
4- E = D -1
5- For the special case of Serine, we find, in addition to the above, that: A + F = B + E = C + D, and C = B-1.

The above numerical laws are summarized in the two tables below.

Sch. 4	Up to Four Codons	Six Codons
	[A, B, C, D]	[A, B, C, D, E, F]
1	[A,B] and [C, D] belong to different triplet groups	[E, F], [A,B] and [C, D] belong to different triplet groups (except for Leucine, where [A, B] and [E, F] comes from the same triplet.)
2	C = B-1 and D = A -1.	E = D -1 (and C = B -1 and A = D-1 for Serine).
3	A+B = C+D.	A+C = B+D and E+C = F+D
4		A+F = B+E (A+F = B+E = C+D for Serine)

(Actually, conditions 3 of the 6-based amino acids automatically determines condition 4 by simple arithmetical mutilations.)

There is one interesting insight to deduce from the above numerical laws, especially number three. This is because I have already encountered it before, on pages 75 and 76 in my book, when I talked about the two types of the numerical Archimedean spirals and how these two types satisfy the same numerical condition 3 of the above tables, but for subtraction instead of addition. And we already know that the DNA itself is a 3-dimensional Archimedean spiral.

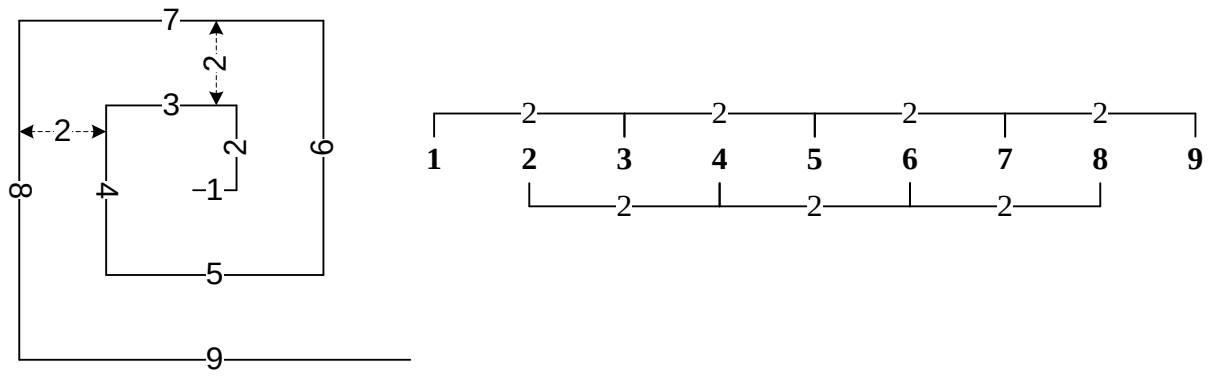


Figure 5: Left: Numerical Archimedean spiral. Right: the numerical condition that Archimedean spirals should satisfy is exactly the same as condition 3: 3-1 = 4-2 = 2, etc.

This spiraling property is inherited from two facts: first, from the fact that A+T = C+G, and secondly from the codonic sequence conditions, which we talked about above, especially that the last nucleotide of each codon in the same amino acid are of the order T, C, A, and G. Even those Amino acids that are made of six codons observe the same law.
Hence, I call this law or condition, condition 3, *the spiraling condition*.

(Could it be that codons coding for amino acids and hence protein are chosen such that to ensure that the spiraling condition is somehow observed, and consequently ensure the helical structure of the whole DNA? Maybe. However, we need to move to more complex systems, like proteins, to verify this claim.)

An Insight into the Distribution of Codons

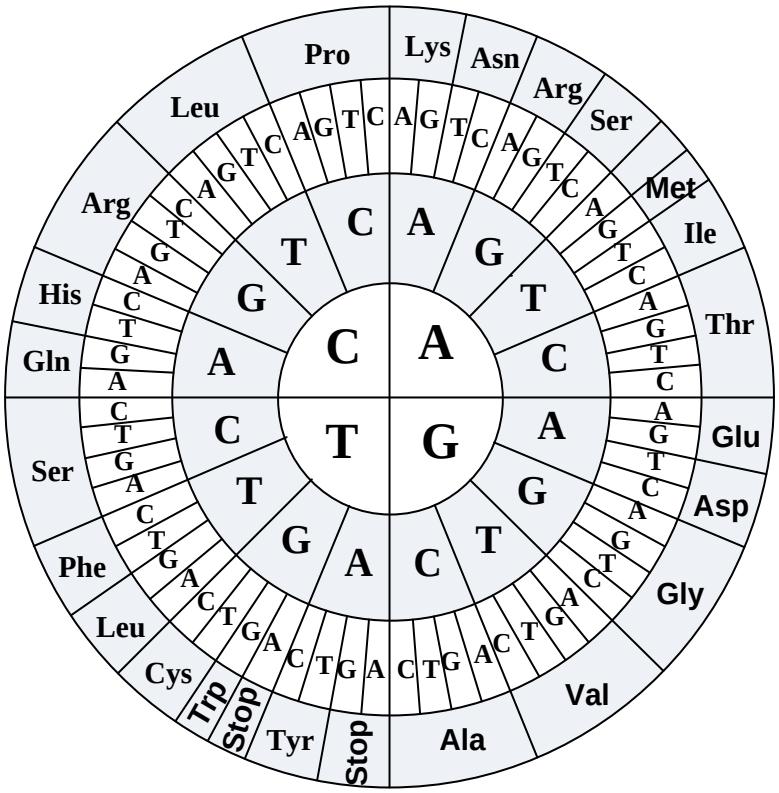
This chapter, in my opinion, is the most important one. As using the amino acids numerical representation we will be able to explain almost all of the codonic distribution within the various codonic families of the

amino acids, proving, and without a doubt; *that numbers resides at the core of the design of the amino acids*, along with the bizarre implication that this numerical design is *based on the treatment of the codons of each codonic family as one whole unit and not separate entities*.

So I start by listing all of the amino acids based on the number of their codons.

Amino Acid	DNA Codons: Sense Strand	Number of Nucleotides	1	2	3	4	5	6	Sum	*D*-Sum
Methionine	ATG	3	1						1	1
Stop Codons	TAA	3	4						4	4
	TAG	3	1						1	1
	TGA	3	1						1	1
Tryptophan	TGG	3	7						7	7
Phenylalanine	TTT, TTC	6	9	3					12	3
Cysteine	TGT, TGC	6	8	2					10	1
Tyrosine	TAT, TAC	6	2	5					7	7
Glutamine	CAA, CAG	6	7	4					11	2
Asparagine	AAT, AAC	6	4	7					11	2
Histidine	CAT, CAC	6	5	8					13	4
Glutamic Acid	GAA, GAG	6	3	9					12	3
Aspartic Acid	GAT, GAC	6	1	4					5	5
Lysine	AAA, AAG	6	6	3					9	9
Isoleucine	ATT, ATC, ATA	9	2	5	4				11	2
Valine	GTT, GTC, GTA, GTG	12	8	2	1	7			18	9
Alanine	GCT, GCC, GCA, GCG	12	2	5	4	1			12	3
Glycine	GGT, GGC, GGA, GGG	12	7	1	9	6			23	5
Proline	CCT, CCC, CCA, CCG	12	6	9	8	5			28	1
Threonine	ACT, ACC, ACA, ACG	12	5	8	7	4			24	6
Serine	TCT, TCC, TCA, TCG, AGT, AGC	18	3	6	5	2	1	4	21	3
Leucine	TTA, TTG, CTT, CTC, CTA, CTG	18	2	8	3	6	5	2	26	8
Arginine	AGA, AGG, CGT, CGC, CGA, CGG	18	3	9	2	5	4	1	24	6

The Codonic families can be also represented in the following circular fashion.



All the amino acids are made from *nine basic codon pairs* divided into three groups, with each belonging to one triplet group, as shown in the table below.

[2, 5, 8]	[1, 4, 7]	[3, 6, 9]
[2, 5]	[1, 4]	[3, 6]
[2, 8]	[1, 7]	[3, 9]
[8, 5]	[7, 4]	[6, 9]

When we take the *D*-sum of the number in each brackets of the above tablet we get the triplet group back again.

[2, 5, 8]	[1, 4, 7]	[3, 6, 9]
7	5	9
1	8	3
4	2	6

There is no apparent reason why this should be the case; why there should be exactly nine basic groups that have a *D*-sum generating numbers from 1 to 9 without any omission or repetition.

There are several fundamental questions that come naturally to one's mind after pondering a bit on the design of the amino acids.
The most obvious one is; why do the codons distribute themselves in such a manner; four for valine, three for isoleucine, etc? Why not have sixteen 4-based codons amino acids only?

Well, it turns out that the answer, or at least the major part of it, lies in the digital root.

To understand how, let us start looking at some amino acids that don't have 4 codons in their codonic family, and see what will happen if they had 4 codons.
Let us start with the closest one to four codons, Isoleucine.
Isoleucine have the following codonic family: ATT, ATC, ATA, with the following values [2, 5, 4] respectively. Now based on the sequence rule we mentioned above, the last codon isoleucine should have, if it to be based on four codons, is ATG. This makes the complete numerical sequence of the codons as follows: [2, 5, 4, 1].
However, this specific numerical code has already been taken by another amino acid, namely alanine.
But does this make any difference?
Well, it turned out it does, as we will discover later on that:

No two different codonic families share the same exact numerical pattern.

This I realized because, when checking for all the non-four-based amino acids, if I to complete them to four, they will always generate the same number sequence of another already existing 4-based amino acid (with the exception of Lysine and Asparagine, which I tackle later on).

In the table below, I recombine all the non-4-based amino acids into four based one and show how they will produce numbers matching other already 4-based amino acids.

Recombined Non-4-Based Amino Acids	Already 4-Based Amino Acids with the Same Numbers
Isoleucine + Methionine [2, 5, 4, 1]	Alanine [2, 5, 4, 1]
Cysteine + Tryptophan+ Stop [8, 2, 1, 7]	Valine [8, 2, 1, 7]
Tyrosine + Stop + Stop [2, 5, 4, 1]	Alanine [2, 5, 4, 1]
Histidine + Glutamine [5, 8, 7, 4]	Threonine [5, 8, 7, 4]
Aspartic Acid + Glutamic Acid [1, 4, 3, 9]	[1, 4] (from Serine) + [3, 9] (from Arginine)

This left us with three combinations which doesn't seem to have a reason to be broken down, (with another question of why the 6-based amino acids existed at all). They are: [3, 9, 1, 4], [6, 3, 4, 7], and [9, 3, 2, 8])

Let us try to solve this mystery by looking first at those of the 6-based amino acids.

6-based Amino Acid	Original 4 Codons	Extra 2 codons
Serine [TCT, TCC, TCA, TCG, AGT, AGC]	[TCT, TCC, TCA, TCG] [3, 6, 5, 2]	[AGT, AGC] [1, 4]
Leucine [TTA, TTG, CTT, CTC, CTA, CTG]	[CTT, CTC, CTA, CTG] [3, 6, 5, 2]	[TTA, TTG] [2, 8]
Arginine [AGA, AGG, CGT, CGC, CGA, CGG]	[CGT, CGC, CGA, CGG] [2, 5, 4, 1]	[AGA, AGG] [3, 9]

As shown from the above table, serine and leucine share the same original 4-based code, hence one of them should have been broken down into two. However they didn't. This has to do with the fact that if one of them is to broken down, it will be [3, 6] and [5, 2]. However [3, 6] is already occupied by Lysine. Consequently, and as none of them can be broken down, their only choice is to add more instead. (But why both of them decided to add two instead of one of them, I still don't know.) Hence, they both decided to add another two to each other, [AGT, AGC] and [TTA, TTG]. Now [AGT, AGC] already exists as a separate 2-based codon as we saw above. The other half went to arginine, which also decided to become a 6-based amino acid and for reasonable causes. This is because its original 4-based code is the same as alanine, and if it were to be broken down, it will regenerate Tyrosine again. Hence, its only option is either to lose one codon (which is not an option as isoleucine has already done that), or to become a 6-based codon, which it did. (This also could be the reason why isoleucine has this weird three codonic composition. This is because breaking into two, two-codon based ones, will not help him either, as one of them will have the same numbers of tyrosine. So losing one was the best option.)

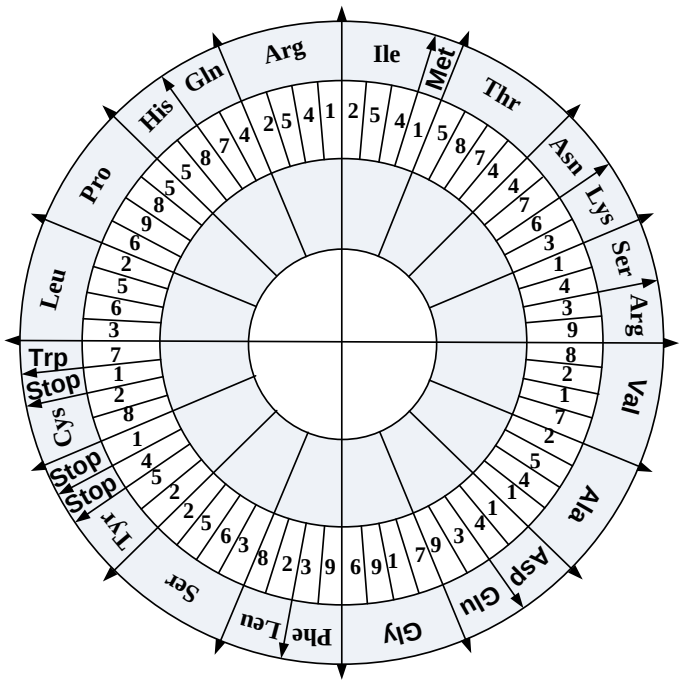
Leucine, on the other hand, once decided to became 6-based amino acid, had no choice but to snap [2, 8] from the [9, 3, 2, 8] of the [TTT, TTC, TTA, TTG] code, leaving [9, 3] by itself. This is because [2, 8] is the only combination of all of the remaining 2-based amino acids that will satisfy the above numerical laws for the 6-based amino acids, especially law 3, the spiraling condition. (Phenylalanine [9, 3] will satisfy the law also, however, then, we would have had two **same** triplet groups one after the other **[9, 3]**, **[9, 6],** [2, 8], which is not allowed, or not preferable, based on the codonic laws.) Adding [7], tryptophan, will not satisfy the spiraling condition as well, neither for leucine, nor for serine or arginine. (Maybe this is why we don't have 5-based amino acids.)

But why didn't isoleucine add another two to its structure instead of losing one? Actually, there are two 2-based amino acids that, when combined with it, will form an amino acid that will not violate any of the

numerical laws above. The two sequences that can, theoretically speaking, work out are [**6, 3**, 2, 5, 4, 1] and [2, 5, 4, 1, **9, 3**]. However, these two codons correspond to [6, 3] (lysine) and [9, 3] (phenylalanine), and both of them are the only ones coding for their corresponding amino acids, and hence cannot be dispensed. Hence, isoleucine's only option was to lose a codon; can't lose two, can't gain two.

Nevertheless, and whatever the real answer may be, the fact remains; none of the above amino acids violate the second numerical law we just discovered; *that no two amino acids should share the same numerical sequence*, which is very intriguing indeed. It is as if this numerical sequence is a signature or fingerprint of the amino acids, in biology as well as in the digital root world. As if there is something that tell the codons about their numerical signature and dictates on them how to behave, as a one unified entity. I mean, if it were only for the first numerical laws, then the whole above study will not make any sense anymore; because the codons, presumably, do not act together. However, the results we just found in this section are very fundamental and conclusive and in no way could have come by chance. But these results make no sense if the codons of each amino acids are not considered as a whole. Hence, we must be missing something when we look at codons as individual coding entities. We miss the whole picture.

Below I show another circular diagram of the various codonic families but plotted in respect to their numbers now instead of their codons sequence



The rest of the 2-codon based amino acids that have no obvious numerical reason to be broken down from 4-codon based system can be explained on the basis that it was necessary step to accommodate for the rest of the amino acids. In other words I have **two** 4-codon based groups but I still have **four** amino acids to go. Consequently, each 4-codon group has to be split into two.

(In fact, one may start with 2 codons for each of the 20 amino acids, and then go on distributing the rest of the combinations between them, along with the stop-start codons, in such a way as to observe the numerical conditions talked about above.)

The table shown below list the nine basic 4-based groups and their corresponding amino acids.

	2541	**5874**	**4763**	**1439**	**8217**	**7196**	**9328**	**3652**	**6985**
	Alanine	Threonine	Asparagine Lysine	+Serine +Arginine	Valine	Glycine	Phenylalanine +Leucine	Serine+	Proline
	Arginine+	Histidine Glutamine		Aspartic Acid Glutamic Acid	Cysteine Stop Tryptophan			Leucine+	
	Isoleucine Methionine								
	Tyrosine 2×Stop								
Total	5	3	2	4	3	1	2	2	1

Interestingly, when we take the digital root of these 4-based groups, we find that they produces numbers 1to 9 again, as shown below.

Group	2541	5874	4763	1439	8217	7196	9328	3652	6985
D-sum	**3**	**6**	**2**	**8**	**9**	**5**	**4**	**7**	**1**

If we to accept that the 2-based basic group, talked about earlier, produced numbers 1-9 by chance, and that their 9-ish number and distribution between the three basic triplet is also by a pure chance, how can we explain this; that the 4-based groups are exactly nine also and produced 1-9 again? If we had 8269 instead of 8217, for example, we would had another 9 instead of 1, and the 1-9 order would have been broken.

Another fundamental question that one may think of is, why do I have one start codon and three stop codons?

To answer this question, let suppose first that we have only one stop codon, let say TAA. This should have worked well, except for the case when it comes after one of the codons of tyrosine, namely [TAC], because then it will complete the sequence in the correct sequence condition we talked above earlier on, and their combination will look as if it belongs to some new amino acid [TAC, TAA], a one that should not exist. Hence a different stop codon should be used, TAG, or even better TGA, as TAG is from the same codonic family of tyrosine [TAT, TAC].
For the same reasons, TGA will not work with cysteine, which has the code of [TGT, TGC]. TAA will work fine to stop this code. (Now, this all a preliminary theoretical speculation from my behalf. It definitely need to be verified)

Numerically, TGA and TAG cannot be used with isoleucine because the numerical pattern will then be [2, 5, 4, 1] which is already occupied by alanine. Consequently, TAA should be used instead.
Moreover, for tryptophan, TGA is not preferable, as it is from the same codonic family. TAA cannot also be used, as it will make the pattern [7, 4], which is already included in Glutamine. Hence, TAG is the perfect choice.

Stop Codon	Amino Acids Cannot Work With	Amino Acids Must Work With
TAG	Isoleucine	Tryptophan
TGA	Cysteine, Isoleucine	Tyrosine
TAA	Tyrosine, Tryptophan	Isoleucine

So, now we know, theoretically at least, why we need these three stop codons, TAA, TAG, and TGA.

For the start codon, on the other hand, the mystery is easier to solve; putting ATG at the beginning will not match any amino acid sequence, neither chemically nor numerically, therefore there is no risk of being confused as a new amino acid. (Of course ATG belongs to the family of Isoleucine, never the less, putting it at the front of this family will not satisfy the codonic sequence condition.)

All the above, I believe, establishes, and without a doubt, the fundamental role numbers play, not only in the abstract world of numerology, but deep within the complexity of the living biological systems as well. As we have already seen how codons split or joined not only based on their chemical codonic structure, but more importantly, based on the numbers assigned to them and on the numerical laws they are governed with.

All the numerical laws we have discovered up till now may very well be extended to explain other observations and phenomenon within the world of genes and DNA, such as why a certain gene become functional, and consequently be copied, and why not, and why a certain hereditary genetic trait is chosen over another one, etc. And also why a certain codon of some amino acid is used in one gene and why another codon of the same amino acid used in a different gene. It could very well be that their numerical values, and how they are incorporated within the final code, while observing the above numerical laws, is what determine how the individual genes behave with the DNA code. Verifying this definitely requires a lot of further research. A computer coding will definitely help a lot.

I think we are now in a position to answer some of the most fundamental questions about DNA and the amino acids.

But, before we proceed, there is one point I would like to address. We are always accustomed to find answers to fundamental questions through math: if 1+1 = 2, then that will do it. However, when we move into the living realm of nature, we should expect that answers might not come solely (or entirely) from a mathematical rezoning anymore; they may come from a meaningful or reason-ful logic instead.
Some of the questions may be answered directly from within the same level of the questions; from the same geometrical or numerical patterns of the amino acids for example. Others may require us to go to a higher level, like the exact chemical composition of the actual amino acids or from the level proteins, etc. Some questions, on the other hand, may not be answered at all, not until we reach the final and most important level, the level of the complete human being; because this design is what it takes to have a fully perfect human being in the way we have it. There is no simpler answer.

Nevertheless, many of the fundamental questions we encountered above do have numerical and mathematical solutions as we saw. Below I list the most important of these questions along with reasonable answers (when I have them of course).

1- Why do codons come in groups of three nucleotides?

There is a numerical reasoning for this as well as a numerological or meaningful one.
Numerically, this because the only way to join the four nucleotides such that it will create a coding pattern equal, or bigger, to the number of amino acids available, which is 20+, is by grouping them as three, three… this is because a code built only on nucleotides joined two, two, will only generate 16

possibilities, which is less that the required 20+ amino acids. And four, four, will generate an abundance of them, much more than needed. Hence, triplet of nucleotides is the best choice.

Numerologically, this is because it takes three to specify an object, to make it complete and give it identity. We need three elements to define an atom; electrons, protons and neutrons. A human being, as well as all living creatures, is defined by his body, soul and spirit, etc. And so is the case with codons. Like a Borromean Knot, you need the three rings to keep the knot unified; once a ring is gone, the whole structure will collapse.

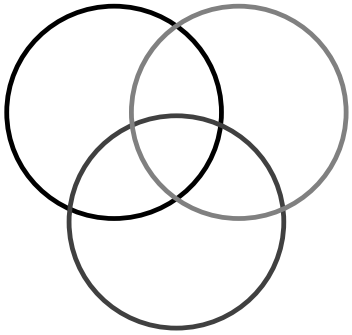


Figure 6: Borromean knot made of three rings interlocked with each other such that of one ring removed the other two will fall apart.

2- Why do amino acids exist with these codonic families? Why didn't we simply have 16 4-codons based amino acids? Why don't we have five codons-based amino acids? And why did the 6-codnos based amino acids formed?

Why didn't we simply have 16 4-codons based amino acids? Well, I don't have an answer to this one. At least no now. This is how we were created; we needed 20 amino acids + 3 for some reason. However, I believe that we have answers to the rest of the questions, may be not entirely, but still, on the overall, I do believe that the numerical laws and patterns we discovered above were able to explain why the majority of the amino acids' codonic families exist the way they do. Other laws may also be discovered in the future, laws that will solve for all the remaining question and anomalies.

3- How does the amino acid know about the numerical value of the codons, let along the digital root?

Most probably through geometry, or maybe through their electromagnetic field. I still have no definite answer. But it is definitely one of the most fundamental questions that cries for an answer.

There are some other question one can ask, such as why glycine is the only one that is not chiral, etc. but again, these question may need to be answered one, or several, levels higher than the level of the nucleotides or codons, which will be tackled later on.

However, the most enigmatic question, for me at least, is: why does the unified treatment of codons matter? Why do we understand more about the codons distribution between the amino acids when we treat them as one block, when they clearly don't work together in the coding process? This I believe is a very fundamental question that needs to be understood if we to be able to unravel some of the mysteries of the DNA.

* * *