Advanced Techniques in Biology & Medicine

Bozegha, Adv Tech Biol Med 2017, 5:4
DOI: 10.4172/2379-1764.1000245

Research Article Open Access

The 64-Triplet Genetic Code Structure Revisited and Refuted from Combinatorial

WB Bozegha*

Numeration Science Literature, Development Research Project, Yenagoa, Bayelsa State, Nigeria

Abstract

Statement of the problem: When molecular biologists observed in 1953 that the sequence of the DNA four bases in the nucleus of a cell influenced the sequence of the twenty amino acids of protein in the cytoplasm, they desired to find a code to account for the correlation, and eventually had the 64 triplet genetic code in 1954 from a mathematician, which is currently in use but not flawless.

Methodology and theoretical orientation: The said observation is seen as a natural example of an input/output system, in which the input is the DNA four bases and the output is the sequence of 24 permutations of the four DNA bases constituting the genetic code in the cytoplasm. A combinatorial input/output multiplicative replication system armed with basic permutations computation schemes is now available to produce permutations systematically, such as Square Kinematics Scheme and Successive Collateral Posting Scheme used.

Findings: A 24 quadruplet genetic code was produced by each of the two methods with an input set of the DNA four bases. It is shown in the successive collateral posting method that the 64 triplets comprise 40 non-permutations and 24 permutations. The 40 non-permutation triplets are crossed out leaving 24 permutation triplets which are undersized and therefore unqualified to represent the genetic code output sequence from an input set of 4.

Conclusion and significance: The 24 quadruplet genetic code is a breakthrough in the Molecular Biologists' search for a code following their observation which ended up with the 64-triplet genetic code that has no combinatorially valid code word being triplets, instead of quadruplets.

Recommendations: These are made towards effective publicity of the new 24 quadruplet genetic code to attract experimental experts to spell it to win adoption in coding application in protein studies for a desired relief to stakeholders in genetics.

Keywords: 64-triplet genetic code; Molecular biology; Cytoplasm; Protein synthesis studies

Introduction

The 64-triplet genetic code structure that begged for revisit is immature and a mixture of 24 permutation and 40 non-permutation triplet codons derived in 1954 by the base-4 neo-digibreed indirect method with input set of RNA four bases A, U, G, C (Adenine, Uracil, Guanine, Cytosine). The revisit is from combinatorial perspective, because the task of deriving a genetic code structure from the given four nucleotide bases is in the province of computational combinatorics and borders on the generation of permutations of 4 from 4 (Figure 1).

The objectives of this work are four-fold, given the challenge to derive afresh a code from the four RNA bases, A, U, G, C (Adenine, Uracil, Guanine, Cytosine) to account for the correlation between them and twenty amino acids of protein as sequel to molecular biologists' observation in the early 1950s, that the sequence of the four nucleotide bases A, U, G, C in the nucleus of a cell influenced the sequence of the 20 amino acids of protein in the surrounding cytoplasm in the cell, because the one of 64-triplet structure based on 4^3 derived in 1954 in response to the scientists' quest and spelt thereafter and adopted since 1968 in coding application in protein synthesis studies is bedeviled with irregularities which are only widely discussed in genetics literature, but without remedy to date, hence the first objective is to derive a genetic code without any irregularity using the same material and indirect method used in 1954 by the authors of the current one now in disfavour. The second objective is to use a direct method to produce the irregularity-free genetic code for confirmation of validity. The third one is to show why the 64-triplet genetic code structure is bedeviled with the well-known irregularities. Lastly, the fourth objective is to highlight the combinatorial and molecular biological merits of the new genetic code structure of 24 quadruplets and thereby attract spelling experts to spell it in order to render it fit for adoption for coding application in protein synthesis studies, in the event of the publication of this work as per this attempt, and that to the relief of all stakeholders worldwide.

Materials and Methods

Materials

The materials consist of the RNA four bases in the sequence of A, U, G, C (Adenine, Uracil, Guanine, Cytosine) as transcription from the DNA four bases in the order of A, T, G, C (Adenine, Thymine, Guanine, Cytosine) as carried in a particular rung of the double helix and are used as input set of 4 bases in the multiplicative replication input/output system in computational combinatorics developed by this author in the 1990's.

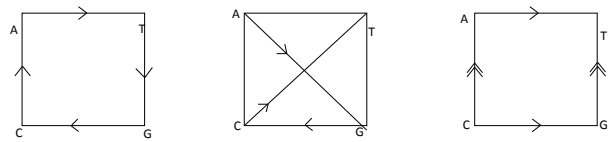
Methods

The methods are two, namely: *base-four neo-digibreed* indirect

***Corresponding author:** WB Bozegha, Numeration Science Literature, Development Research Project, No. 9 Otiotio Road, Yenagoa, Bayelsa State, Nigeria, Tel: +2348068029860; E-mail: wbbozegha@yahoo.com

Received September 05, 2017; **Accepted** September 29, 2017; **Published** October 06, 2017

Citation: Bozegha WB (2017) The PE/PPE Multigene Family of Mycobacteria and TB Vaccines. Adv Tech Biol Med 5: 231. doi: 10.4172/2379-1764.1000245



(a) Sides deployment for production of 8 quadruplet permutations with input set ATGC

(b) Diagonals deployment for production of 8 quadruplet permutations with input set ATGC

(c) Parallels deployment for production of 8 quadruplet permutations with input set ATGC

Figure 1: Square kinematics mixing scheme for computing 4 from 4 permutations by view mixing.

method, Chart 1 and *Square kinematics* direct method, Chart 2 under Figure 2.

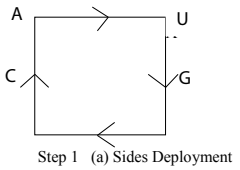
First, base-four neo-digibreed indirect method uses an input set of the RNA four bases in the linear sequence of A,U,G,C involving successive collateral posting (SCP) in *Fati's Geotropic Frame* to the limit of 4^4 where 256 quadruplets are produced. This product is a mixture of 24 permutation quadruplets (crops) bearing no repetition(s) of letters, and 232 non-permutation quadruplets (weeds) bearing repeated letters as carried in Chart 1 in lines 22-85. The 232 non-permutation quadruplets are crossed out as weeds to leave a residue of 24 permutation quadruplets as crops for harvesting as the valid codons.

Corridor	Digibreed	Line No.	Output No. of permutations per line per digitality		
	Col1 Col 2 Col 3 Col 4 Input A U G C		Digitality 2	Digitality 3	Digitality 4
	OUTPUT				
A	AA AU AG AC	2	3		
U	UA UU UG UC	3	3		
G	GA GU GG GC	4	3		
C	CA CU CG CC	5	3		
Valid total no. of permute-tion doublets per col.	3 3 3 3		12		
AA	AAA AAU AAG AAC	6	-		
AU	AUA AUU AUG AUC	7		2	
AG	AGA AGU AGG AGC	8		2	
AC	ACA ACU ACG ACC	9		2	
UA	UAA UAU UAG UAC	10		2	
UU	UUA UUU UUG UUC	11		-	
UG	UGA UGU UGG UGC	12		2	
UC	UCA UCU UCG UCC	13		2	
GA	GAA GAU GAG GAC	14		2	
GU	GUA GUU GUG GUC	15		2	
GG	GGA GGU GGG GGC	16		-	
GC	GCA GCU GCG GCC	17		2	
CA	CAA CAU CAG CAC	18		2	
CU	CUA CUU CUG CUC	19		2	
CG	CGA CGU CGG CGC	20		2	
CC	CCA CCU CCG CCC	21		-	
Valid total No. of permutation triplets per col.	6 6 6 6			24	
AAA	AAAA AAAU AAAG AAAC	22			-
AAU	AAUA AAGU AAGG AACG	23			-
AAG	AAGA AAGU AAGG AACG	24			-
AAC	AACA AACU AACG AACC	25			-
AUA	AUAA AUAU AUAG AUAC	26			-
AUU	AUUA AUUU AUUG AUUC	27			-
AUG	AUGA AUGU AUGG AUGC	28			1
AUC	AUCA AUCU AUCG AUCC	29			1
AGA	AGAA AGAU AGAG AGAC	30			-
AGU	AGUA AGUU AGUG AGUC	31			1
AGG	AGGA AGGU AGGG AGGC	32			-
AGC	AGCA AGCU AGCG AGCC	33			1

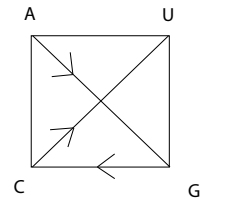
corridor	Digibreed	Line No	Output No. of permutations per line per digitality		
	Col1 Col 2 Col 3 Col 4 Input A U G C		Digitality 2	Digitality 3	Digitality 4
	OUTPUT				
ACA	ACAA ACAU ACAG ACAC	34			-
ACU	ACUA ACUU ACUG ACUC	35			1
ACG	ACGA ACGU ACGG ACGC	36			1
ACC	ACCA ACCU ACCG ACCC	37			-
UAA	UAAA UAAU UAAG UAAC	38			-
UAU	UAUA UAUU UAUG UAUC	39			-
UAG	UAGA UAGU UAGG UAGC	40			1
UAC	UACA UACU UACG UACC	41			1
UUA	UUAA UUAU UUAG UUAC	42			-
UUU	UUUA UUUU UUUG UUUC	43			-
UUG	UUGA UUGU UUGG UUGC	44			-
UUC	UUCA UUCU UUCG UUCC	45			-
UGA	UGAA UGAU UGAG UGAC	46			1
UGU	UGUA UGUU UGUG UGUC	47			-
UGG	UGGA UGGU UGGG UGGC	48			-
UGC	UGCA UGCU UGCG UGCC	49			1
UCA	UCAA UCAU UCAG UCAC	50			1
UCU	UCUA UCUU UCUG UCUC	51			-
UCG	UCGA UCGU UCGG UCGC	52			1
UCC	UCCA UCCU UCCG UCCC	53			-
GAA	GAAA GAAU GAAG GAAC	54			-
GAU	GAUA GAAU GAUG GAUC	55			1
GAG	GAGA GAGU GAGG GAGC	56			-
GAC	GACA GACU GACG GACC	57			1
GUA	GUAA GUAU GUAG GUAC	58			1
GUU	GUUA GUUU GUUG GUUC	59			-
GUG	GUGA GUGU GUGG GUGC	60			-
GUC	GUCA GUCU GUCG GUCC	61			1
GGA	GGAA GGAU GGAG GGAC	62			-
GGU	GGUA GGUU GGUG GGUC	63			-
GGG	GGGA GGGU GGGG GGGC	64			-
GGC	GGCA GGCU GGCG GGCC	65			-
GCA	GCAA GCAU GCAG GCAC	66			1
GCU	GCUA GCUU GCUG GCUC	67			1
GCG	GCGA GCGU GCGG GCGC	68			-
GCC	GCCA GCCU GCCG GCCC	69			-
CAA	CAAA CAAU CAAG CAAC	70			-
CAU	CAUA CAUU CAUG CAUC	71			1
CAG	CAGA CAGU CAGG CAGC	72			1
CAC	CACA CACU CACG CACC	73			-

Corridor	Digibreed	Line No.	Output No. of permutations per line per digitality		
	Col1 Col 2 Col 3 Col 4 Input A U G C		Digitality 2	Digitality 3	Digitality 4
	OUTPUT				
CUA	CUAA CUAU CUAG CUAC	74			1
CUU	CUUA CUUU CUUG CUUC	75			-
CUG	CUGA CUGU CUGG CUGC	76			1
CUC	CUCA CUCU CUCG CUCC	77			-
CGA	CGAA CGAU CGAG CGAC	78			-
CGU	CGUA CGUU CGUG CCGC	79			1
CGG	CCGA CCGU CCGG CCGC	80			1
CGC	CCCA CCCU CCCG CCCC	81			-
CCA	CCAA CCAU CCAG CCAC	82			-
CCU	CCUA CCGU CCGG CCGC	83			-
CCG	CCGA CCGU CCGG CCGC	84			-
CCC	CCCA CCCU CCCG CCCC	85			-
Valid total no. of permutation quadruplets per Col.	6 6 6 6				24
Valid total no. of permutation quadruplets per digitality	6 6 6 6		12	24	24

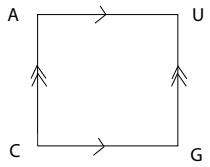
Chart 1: Derivation of 24 quadruplet genetic code structures from the four RNA bases A, U, G, C by successive collateral posting method using base four neo-digibreed and de-isodigitation.
Note: The 24 quadruplet digitisms standing in lines 22 to 85 represent the valid genetic code structure in agreement with combinatorics
Summary of valid permutation codons
Calculated value ${}_4P_4=4! = 4 \times 3 \times 2 \times 1 = 24$ quadruplets
Computed value ${}_4P_4=24$ quadruplets



Step 1 (a) Sides Deployment



Step 2, (b) Diagonals Deployment



Step 3, (c) Parallels Deployment

Chart 3: **Output of $_4P_4$Permutation**

From A	Clockwise	AUGC	Line1
	Fro	CGUA	Line 2
From U	Clockwise	UGCA	Line 3
	Fro	ACGU	Line 4
From G	Clockwise	GCAU	Line 5
	Fro	UACG	Line6
From C	Clockwise	CAUG	Line 7
	Fro	GUAC	Line 8
From A	Clockwise	AGCU	Line 9
	Fro	UCGA	Line 10
From U	Clockwise	UCAG	Line 11
	Fro	GACU	Line 12
From G	Clockwise	GAUC	Line 13
	Fro	CUAG	Line 14
From C	Clockwise	CUGA	Line 15
	Fro	AGUC	Line 16
AU//CG	Horizontals	AUCG	Line 17
	Fro	GCUA	Line 18
UA//GC	Horizontals	UAGC	Line 19
	Fro	CGAU	Line 20
CA//GU	Verticals	CAGU	Line 21
	Fro	UGAC	Line 22
AC//UG	Verticals	ACUG	Line 23
	Fro	GUCA	Line 24

21

Chart 2: Output of 4P4 Permutation.
Summary of valid products: lines 1-24=24 quadruplets
Factorial 4P4=4!=4 × 3 × 2 × 1=24 quadruplets
Production= prediction=24 quadruplets

Direct method, designated as square kinematics: The input set of RNA four bases A, U, G, C are loaded at the corners of the square in clockwise direction as depicted in Figure 3. The loaded square is deployed in three ways as depicted in Figures 3a-3c, to generate 8 combinatorially valid quadruplets per deployment per section using kinematics and view mixing as shown in Chart 3 carrying a genetic code structure of 24 quadruplet codons from lines (1) to (24).

Results

A genetic code structure of 24 permutation quadruplets is presented in Table 1 as a computational reality for the result, being a combinatorial derivation with the two sources or methods stated. The list of 20 amino acids of protein is adapted from the book [1].

Discussion

This discussion is geared to exploring the significance of the results of the work in the context of the four objectives of the revisit set out in the introduction as follows:-

i. Derivation of a genetic code without any irregularity, using the same material and indirect method used in 1954 by the authors of the current 64-triplet code;

ii. Using the same material of RNA four bases A, U, G, C to produce an irregularity-free genetic code structure by a direct method for confirmation of validity;

iii. To show why the 64-triplet genetic code is bedeviled with the much publicized irregularities; and

iv. To highlight the combinatorial and molecular biological merits of the new genetic code structure of 24 quadruplet codons and thereby attract spelling experts to spell it in order to render it fit for adoption in coding application, and that, to the relief of all stakeholders worldwide.

The significance of the result of 24-quadruplet genetic code structure recorded in Figure 4 under item (i) above is, it is concise and precise. The difference in quantum in terms of codewords and base units from 64 codewords to 24 code words, i.e., a reduction to less than half and in base units from 64 × 3 equal to 192 to 24 × 4 equal to 96, showing a reduction of base units involved to exactly half, is clear. The significance here boils down to less labour, besides combinatorial accuracy. Secondly, inspection of Chart 1 lines 6-20 will reveal some code words amongst columns A, U, G, C crossed out, which upon counting would be 40 in number. They are infested with isodigitation, i.e., repetition of digits (bases) and therefore belong to the category of digitism (code words at this instance) known in computational combinatorics as non-permutations. The number of code words surviving the crossing out in the designated region of the chart upon counting would be 24. They are free from isodigitation or base repetition(s) and are said to belong to the category of code words known in combinatorics as permutations. Only permutations are qualified for engagement in the genetic code structure which is a natural example of input/output multiplicative replication

AUGC	INITIAL INPUT SET USING SQUARE KINEMATICS TECHNIQUE (SEE APPENDIX)																							
AUGC 1	AUGC	CGUA	UGCA	ACGU	GCAU	UACG	CAUG	GUAC	AGCU	UCGA	UCAG	GACU	GAUC	CUAG	CUGA	AGUC	ACUG	GCUA	UAGC	CGAU	CAUG	UGAC	ACUG	GUCA
	S 1	S 2	AC 1	AC 2	AC 3	AC 4	AC 5	AC 6	AC 7	AC 8	AC 9	AC 10	AC 11	AC 12	AC 13	AC 14	AC 15	AC 16	AC 17	AC 18	AC 19	AC 20	S 3	S 4
CGUA 2	CGUA	AUGC	GUAC	CAUG	UACG	GCGA	ACGU	UGCA	CUAG	GAUC	GACU	UCAG	UCGA	AGCU	AGUC	CUGA	CGAU	UAGC	GCUA	AUCG	ACUG	GUCA	CAGU	UGAC
	S 2	S 1	AC 6	AC 5	AC 4	AC 3	AC 2	AC 1	AC 12	AC 11	AC 10	AC 9	AC 8	AC 7	AC 14	AC 13	AC 18	AC 17	AC 16	AC 15	S 3	S 4	AC 19	AC 20
UGCA 3	UGCA	ACGU	GCAU	UACG	CAUG	GUAC	AUGC	CGUA	UCAG	GACU	GAUC	CUAG	CUGA	AGUC	AGCU	UCGA	UGAC	CAGU	GUCA	ACUG	AUCG	GCUA	UAGC	CGAU
	AC 1	AC 2	AC 3	AC 4	AC 5	AC 6	S 1	S 2	AC 9	AC 10	AC 11	AC 12	AC 13	AC 14	AC 7	AC 8	AC 20	AC 19	S 4	S 3	AC 15	AC 16	AC 17	AC 18
ACGU 4	ACGU	UGCA	CGUA	AUGC	GUAC	CAUG	UACG	GCAU	AGUC	CUGA	CUAG	GAUC	GACU	UCAG	UCGA	AGCU	ACUG	GUCA	CAGU	UGAC	UAGC	CGAU	AUCG	GCUA
	AC 2	AC 1	S 2	S 1	AC 6	AC 5	AC 4	AC 3	AC 14	AC 13	AC 12	AC 11	AC 10	AC 9	AC 8	AC 7	S 3	S 4	AC 19	AC 20	AC 17	AC 18	AC 15	AC 16
GCAU 5	GCAU	UACG	CAUG	GUAC	AUGC	CGUA	UGCA	ACGU	GAUC	CUAG	CUGA	AGUC	AGCU	UCGA	UCAG	GACU	GCUA	AUCG	CGAU	UAGC	UGAC	CAGU	GUCA	ACUG
	AC 3	AC 4	AC 5	AC 6	S 1	S 2	AC 1	AC 2	AC 11	AC 12	AC 13	AC 14	AC 7	AC 8	AC 9	AC 10	AC 16	AC 15	AC 18	AC 17	AC 20	AC 19	S 4	S 3
UACG 6	UACG	GCAU	AUGC	UGCA	CGUA	AUGC	GUAC	CAUG	UCGA	AGCU	AGUC	CUGA	CUAG	GAUC	GACU	UCAG	UAGC	CGAU	AUCG	GCUA	GUCA	ACUG	UGAC	CAGU
	AC 4	AC 3	AC 2	AC 1	S 2	S 1	AC 6	AC 5	AC 8	AC 7	AC 14	AC 13	AC 12	AC 11	AC 10	AC 9	AC 17	AC 18	AC 15	AC 16	S 4	S 3	AC 20	AC 19

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

CAUG 7	CAUG	GUAC	AUGC	CGUA	UGCA	ACGU	GCAU	UACG	CUGA	AGUC	AGCU	UCGA	UCAG	GACU	GAUC	CUAG	CAGU	UGAC	ACUG	GUCA	GCUA	AUCG	CGAU	UAGC
	AC 5	AC 6	S 1	S 2	AC 1	AC 2	AC 3	AC 4	AC 13	AC 14	AC 7	AC 8	AC 9	AC 10	AC 11	AC 12	AC 19	AC 20	S 3	S 4	AC 16	AC 15	AC 18	AC 17
GUAC 8	GUAC	CAUG	UACG	GCAU	ACGU	UGCA	CGUA	AUGC	GACU	UCAG	UCGA	AGCU	AGUC	CUGA	CUAG	GAUC	GUCA	ACUG	UGAC	CAGU	CGAU	UAGC	GCUA	AUCG
	AC 6	AC 5	AC 4	AC 3	AC 2	AC 1	S 2	S 1	AC 10	AC 9	AC 8	AC 7	AC 14	AC 13	AC 12	AC 11	S 4	S 3	AC 20	AC 19	AC 18	AC 17	AC 16	AC 15
AGCU 9	AGCU	UCGA	GCUA	AUCG	CUAG	GAUC	UAGC	CGAU	ACUG	GUCA	GUAC	CAUG	CAGU	UGCA	UGCA	ACGU	AGUC	CUGA	GACU	UCAG	UACG	GCAU	AUGC	CGUA
	AC 7	AC 8	AC 16	AC 15	AC 12	AC 11	AC 17	AC 18	S 3	S 4	AC 6	AC 5	AC 19	AC 20	AC 1	AC 2	AC 14	AC 13	AC 10	AC 9	AC 4	AC 3	S 1	S 2
UCGA 10	UCGA	AGCU	CGAU	UAGC	GAUC	CUAG	AUCG	GCUA	UGAC	CAGU	CAUG	GUAC	GUCA	ACUG	ACGU	UGCA	UCAG	GACU	CUGA	AGUC	AUGC	CGUA	UACG	GCAU
	AC 8	AC 7	AC 18	AC 17	AC 11	AC 12	AC 15	AC 16	AC 20	AC 19	AC 5	AC 6	S 4	S 3	AC 2	AC 1	AC 9	AC 10	AC 13	AC 14	S 1	S 2	AC 4	AC 3
UCAG 11	UCAG	GACU	CAGU	UGAC	AGUC	CUGA	GUCA	ACUG	UAGC	CGAU	CGUA	AUGC	AUCG	GCUA	GCAU	UACG	UCGA	AGCU	CUAG	GBUC	GUAC	CAUG	UGCA	ACGU
	AC 9	AC 10	AC 19	AC 20	AC 14	AC 13	S 4	S 3	AC 17	AC 18	S 2	S 1	AC 15	AC 16	AC 3	AC 4	AC 8	AC 7	AC 12	AC 11	AC 6	AC 5	AC 1	AC 2
GACU 12	GACU	UCAG	ACUG	GUCA	AGUC	UGAC	CAGU	GCUA	AUCG	AUGC	CGUA	CGAU	UAGC	UACG	UACG	GCAU	GAUC	CUAG	AGCU	ACGA	UGCA	ACGU	GUAC	CAUG
	AC 10	AC 9	S 3	S 4	AC 13	AC 14	AC 20	AC 19	AC 16	AC 15	S 1	S 2	AC 18	AC 17	AC 4	AC 3	AC 11	AC 12	AC 7	AC 8	AC 1	AC 2	AC 6	AC 5

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

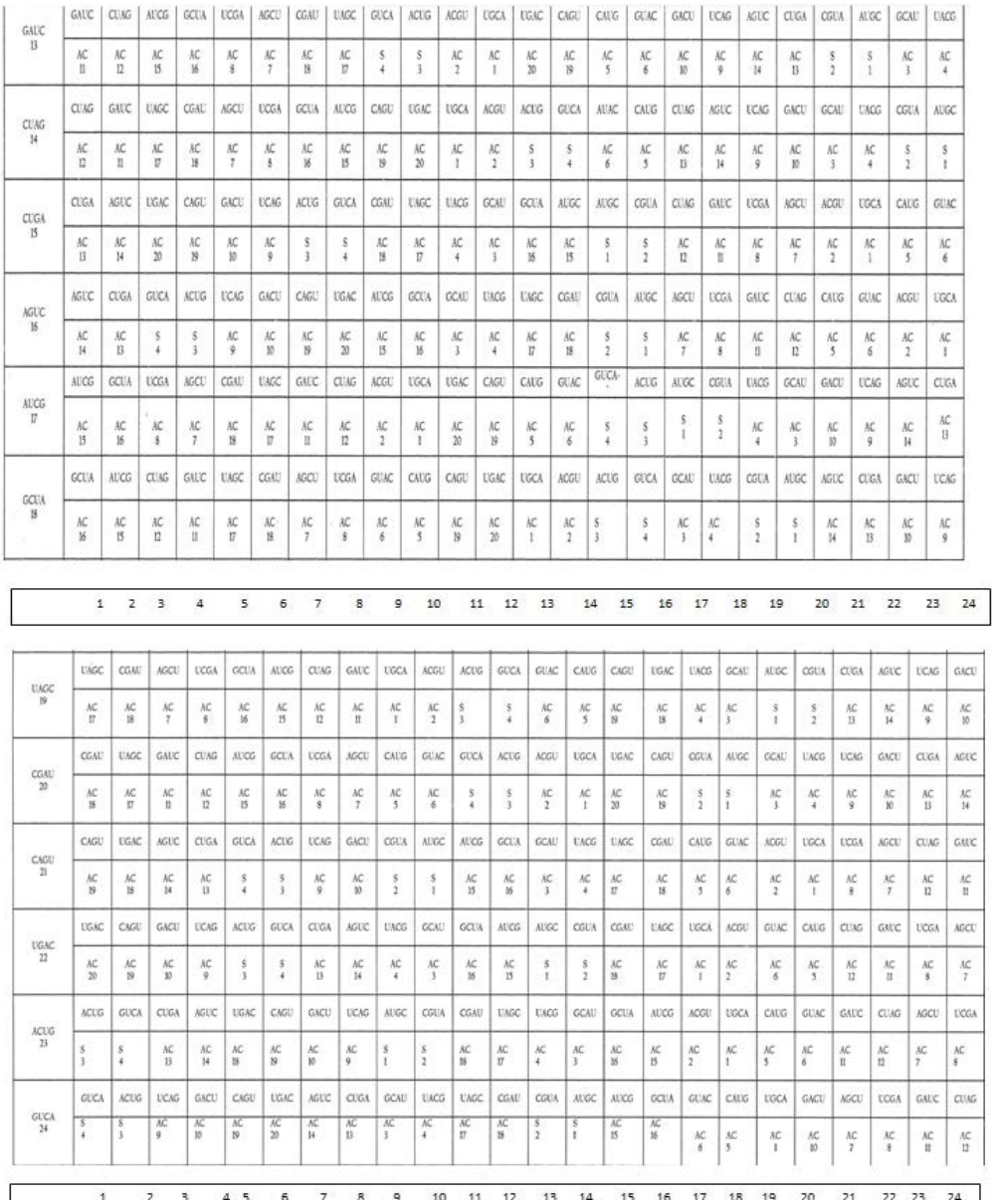


Chart 3: Protein type proliferation and diversification, climax of merits of quadruplets codons.
Key to Chart 3
S1=Signal 1=Place Start signal
S2=Signal 2=Time start signal
S3=Signal 3=Place stop signal
S4=Signal 4=Time stop signal
AC: Amino acid; AC (1)-(2): Amino Acids Numbering 20

system in computational combinatorics. So the distinction made between permutations (crops, bearing no repeated letters) and non-permutations (weeds, bearing some repeated letters) amongst the 64 triplets produced by the indirect method of *successive collateral posting (SCP)* is a remarkable significance in the derivation of the true genetic code structure from the input set of A, U, G, C. The 24 permutation triplets surviving the crossing out or de-isodigitation, nevertheless, are immature and undersize relative to the input set of 4 and are therefore disqualified as code words, hence the chart is continued to the digitality level 4 combination covering lines 22-85 wherein 256 quadruplets are carried. Of these only 24 are permutations (crops) and are standing as residue for harvesting, while 232 are crossed out being non-permutations (weeds). The distinction between permutations and non-permutations has therefore helped to yield 24 valid code words from a list of 256 quadruplets produced by the indirect method of SCP or the base-4 neo-digibreed indirect method of permutation production (Chart 1). Therefore another significance of the result under item (i) is the offer of a reliable indirect method or technique for deriving the true genetic code structure of 24 quadruplets of the status of permutation. If we might consider the results of the work in terms

Input set of RNA four bases	Output permutations of quadruplet codons					Remarks
	Serial no. of 24 quadruplet codons	20 amino acids[a]/4 signals	Amino acids 20/ signals 4 to be specified by codons upon spelling by the experimental experts	Methods		
				(1) Indirect	(2) Direct	
				Base-four neo-digibreed, Ref. Chart 2 lines 22-85	Square Kinematics, Ref. Chart 3 lines 1-24	
		[1]		Genetic Code Sequence	Genetic Code Sequence	Some salient points
AUGC	1	Alanine	YTBD	AUGC	AUGC	(1) Output sequence of permutations per method is unique. (2) All 24 codons per sequence possess integrity and potency. (3) Collinearity between 24 codons and 20 amino acids/4 signals evident. (4) All codons are convertible to equivalents of DNA rungs in base content by the replacement of U by T showing that the genetic code is actually the RNA transcribed from the DNA as intimated [2].
	2	Arginine	"	AUCG	CGUA	
	3	Asparagine	"	AGUC	UGCA	
	4	Aspartic acid	"	AGCU	ACGU	
	5	Cysteine	"	ACUG	UAGC	
	6	Glutamic acid	"	ACGU	UACG	
	7	Glutamine	"	UAGC	CAUG	
	8	Glysine	"	UACG	GUAC	
	9	Histedine	"	UGAC	AGUC	
	10	Isoleucine	"	UGCA	UCGA	
	11	Leucine	"	UCAG	UCAG	
	12	Lysine	"	UCGA	GACU	
	13	Methionine	"	GAUC	GAUC	
	14	Phenylalanine	"	GACU	CAUG	
	15	Proline	"	GUAC	CUGA	
	16	Serine	"	GUCA	AGUC	
	17	Threonine	"	GCAU	AUCG	
	18	Tryptophan	"	GCUA	GCUA	
	19	Tyrosine	"	CAUG	UAGC	
	20	Valine	"	CAGU	CGAU	
	21	Signal 1	"	CUAG	ACUG	
	22	Signal 2	"	CUGA	GUCA	
	23	Signal 3	"	CGAU	CAGU	
	24	Signal 4	"	CGUA	UGAC	
Total	24	24	24	24	24	

Table 1: New genetic code of 24-quadruplet codon structure.
([a]List of 20 amino acids of protein adapted from Figure. 17.4, The World of Cell, p.529 by Becker, Wayne M. (1986). [b] By Jill Wright et al (1988) in their book, Prentice Hall Life Science at page 63 with regard to protein synthesis, where it is stated that the RNA in the ribosomes, along with the RNA sent out from the nucleus directs the production of proteins)
YTBD: Yet To Be Determined; Signal 1: Place Start Signal; Signal 2: Time Start Signal; Signal 3: Place Stop Signal; Signal 4: Time Stop Signal

of *software* and *hardware*, then under item (ii) above, the significance of the result of the work can be identified as the provision of additional direct method or combinatorial technique (hardware) for deriving a correct 24-quadruplet genetic code structure (software) from the same material input set of RNA four bases A, U, G, C with precision as per Chart 2 under Figure 2. Thus the square kinematics technique for direct production of the true genetic code structure is of much significance in molecular biology in the context of protein synthesis studies. With the 24-quadruplet structure confirmed by an independent method, the validity of the derivation result is assured, and it is definitely a breakthrough in the derivation of the combinatorially correct genetic code structure. Under item (iii) above: "To show why the 64-triplet genetic code is bedeviled with the much publicized irregularities" as per Figure 4, the significance of the result of the work (revisit) is the identification and elimination of the forty non-permutation triplets as combinatorial irregularities of the genetic code structure as the combinatorial output of the input set of RNA four bases A, U, G, C, by the indirect base-4 neo-digibreed method.

Combinatorial error is detected in the derivation of the 64-triplet genetic code structure from the four RNA bases A, U, G, C, as shown in Figure 5. The source of the error is traced to the wrongful interpretation of selections for permutation as neo-digibreed population formula (b^d), where b is base-strength or input set and d is the digitality required, instead of permutation factorial complements for set (n) and selection (r) given by (r) from (n), i.e., $nPr=\frac{n!}{(n-r)!}$ as depicted in Figure 4: "*The parted ways of right and wrong at selections for permutation in the derivation of genetic code from input set of four RNA bases in box-graphics*".

The significance of greater importance is that the new 24-quadruplet genetic code structure is a corrective measure freed of all the irregularities associated with the 64-triplet structure and doubling as a refutal to the degenerate code to the end that there is no room again for hypotheses of the kind of the Wobble [3] phenomenon. The derivative is now a true copy of the genetic code engaged in protein synthesis in plants and animals, flawless, since creation and upholding the inerrancy of Nature all the way.

Lastly, under item (iv) above, on highlighting the combinatorial and molecular biological merits of the new genetic code structure of 24 quadruplets of permutation status, the significance of the result of the work (revisit) in molecular biology as conveyed in Chart 3 "*Protein type proliferation and diversification, climax of merits of quadruplet codons*" is two-fold. Firstly, the material is easily adapted to demonstrate the working of the observation in the early 1950s by molecular biologists that, the sequence of the four RNA bases A, U, G, C, in the nucleus of a cell, influenced the sequence of the 20 amino acids of protein in the surrounding cytoplasm of the cell as per Chart 3 in which Column 1 bearing the input set A, U, G, C, at the top and the output of 24 quadruplets is meant to represent the nucleus of a cell, while Column 2 carrying the various output sequences of the input sets in numbered rows 1-24 is meant to represent the cytoplasm of the cell. There are two subrows per numbered horizontal chamber in Column 2. The upper subrow carries the output sequence of the genetic code structure corresponding to the input set in the adjoining chamber in Column 1. The lower subrow per numbered horizontal chamber in Column 2 carries the related protein type with four interspersed codon-sized empty spaces left by four signals 1-4 for time and place based start/stop controls in protein

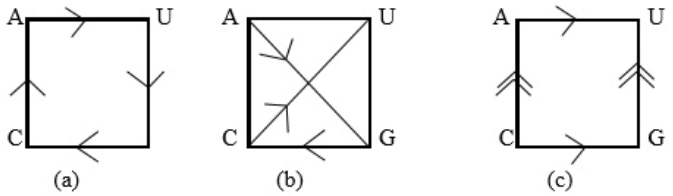


Figure 2: Square kinematics technique for generating permutations of 4 from 4: input set AUGC (**a:** Sides deployment; **b:** Diagonals deployment; **c:** Parallels deployment).

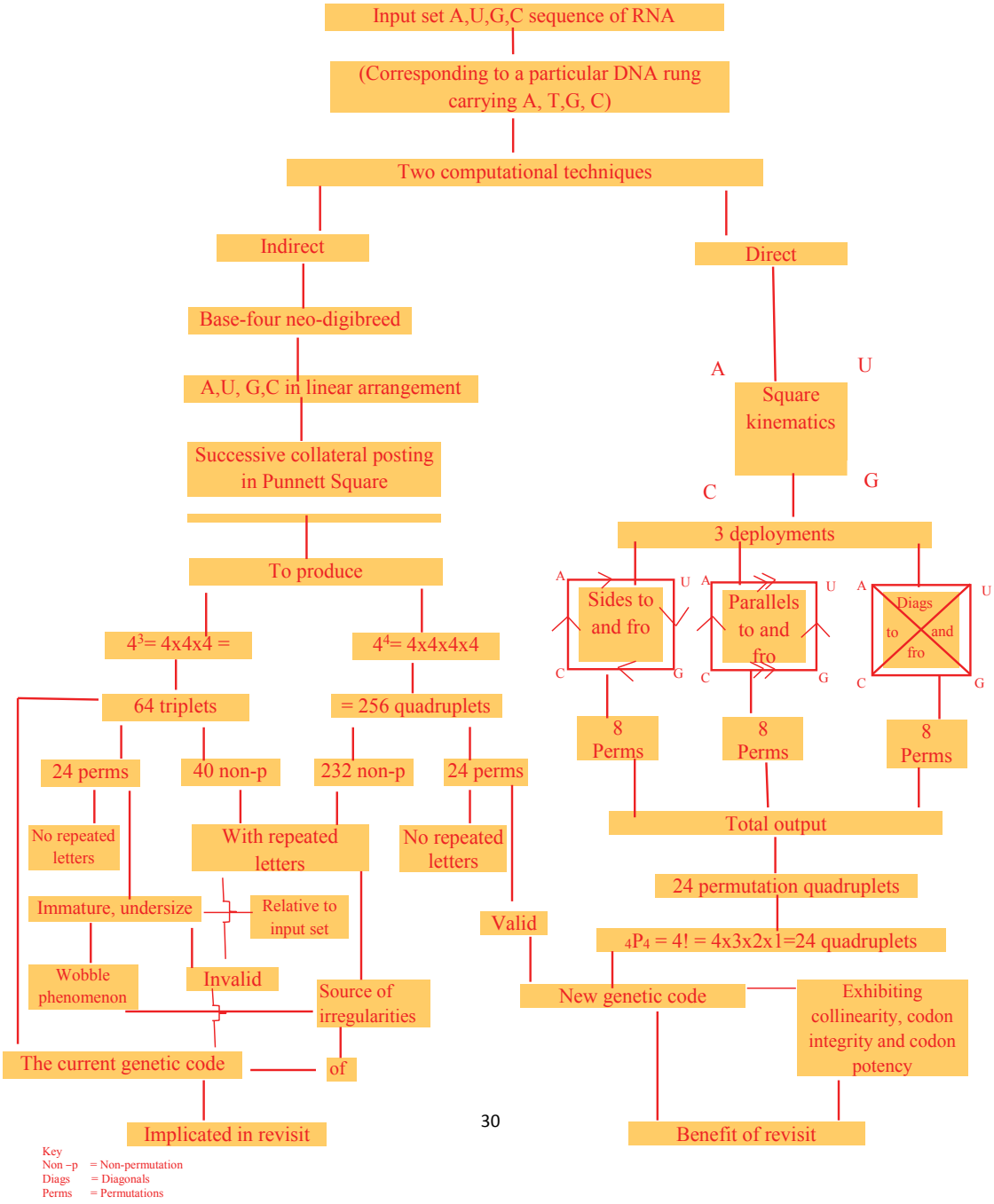


Figure 3: Genetic code derivation from four nucleotide bases A, U, G, C in box-graphics.

synthesis. The configuration of the chart ideally consists of odd rows to the right and even rows to the left of Column 1 in order to represent the neighbourhood of nucleus and cytoplasm in the cell, but it is altered to what is seen in the chart, because of space constraint.

Secondly, Chart 3 carrying 24 protein types (in the lower sub rows) is an illustration of ABC of protein type proliferation and diversification. Chart 3 is evidence of the efficiency of the new genetic code of 24-quadruplet structure which can boast of collinearity (Chart 3, Column 2). Two other molecular biological merits are that each quadruplet codon is identifiable with a specific amino acid or control signal and is capable of reproducing the entire genetic code structure when deployed as input set. Both functions are exercisable in furtherance of protein synthesis. Two more molecular biological merits of the new genetic code structure hinged on codon integrity are the workability of both Chargaff's rules of A=T or U and G=C [4] and Watson-Crick's base pairing of A/T or U and G/C [5] as illustrated in Diagram 1 captioned "*New Genetic Code Structure in Dendritic Dichotomization*", where 2(A=C=12) lines and 2(G=C=12) lines per genetic code sequence of 24 quadruplet codons bearing 96 nucleotide base units in four kinds in keeping with Chargaff's rules. Also base

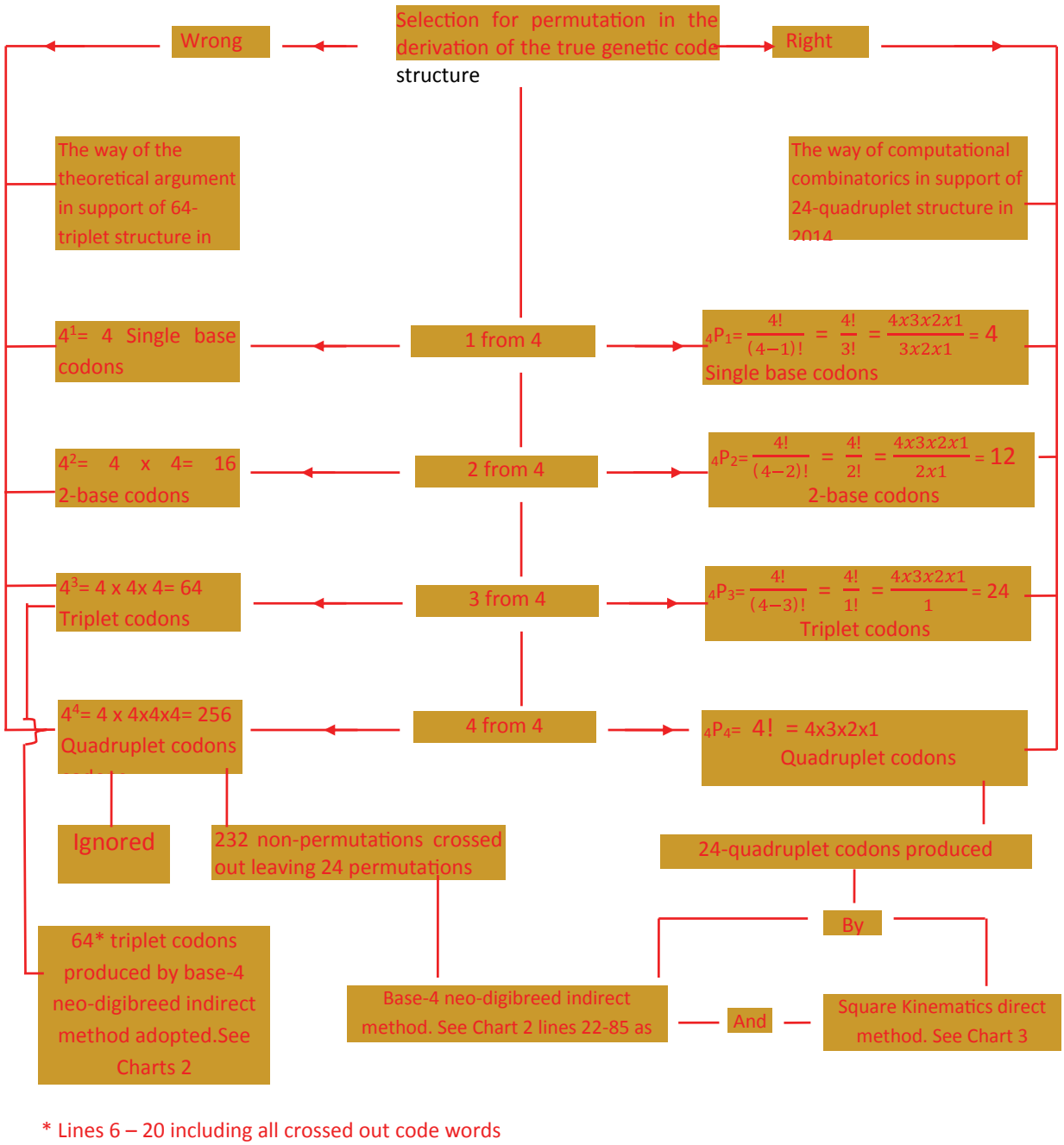


Figure 4: The parted ways of right and wrong at selections for permutation in the derivation of genetic code from input set of 4 RNA bases in box-graphics.

pairing of 2 (A/U × 12) lines and 2 (G/C × 12) lines per genetic code sequence of 24 quadruplets codons featuring 96 nucleotide base units in four kinds in conformity with Watson-Crick's rules is apparent. Yet another remarkable molecular biological merit of the new genetic code structure is that it represents any portion of the DNA double helix comprising 24 consecutive rungs in base-content as depicted in Chart 3 Column 1, when U is replaced by T. The combinatorial merits are exemplified by codon integrity, codon potency, and codon deployability as input set in the input/out multiplicative replication system of computational combinatorics, all owed to permutation status as shown in Figure 6, titled, *Combinatorial essentials of theoretical plant flow-chart for protein synthesis and primary folding of protein type in box-graphics.* "It is also noteworthy that the work has asked and answered a fundamental question in protein synthesis, "Why the genetic code in protein synthesis?" as per Figure 6 parts 1 and 2. Hence the genetic code option to justify its adoption in nature must be the workforce of 24 quadruplet codons that it is with collinearity and repeatability for continuity and sustainability that can match the unending task of protein synthesis in plants and animals since creation.

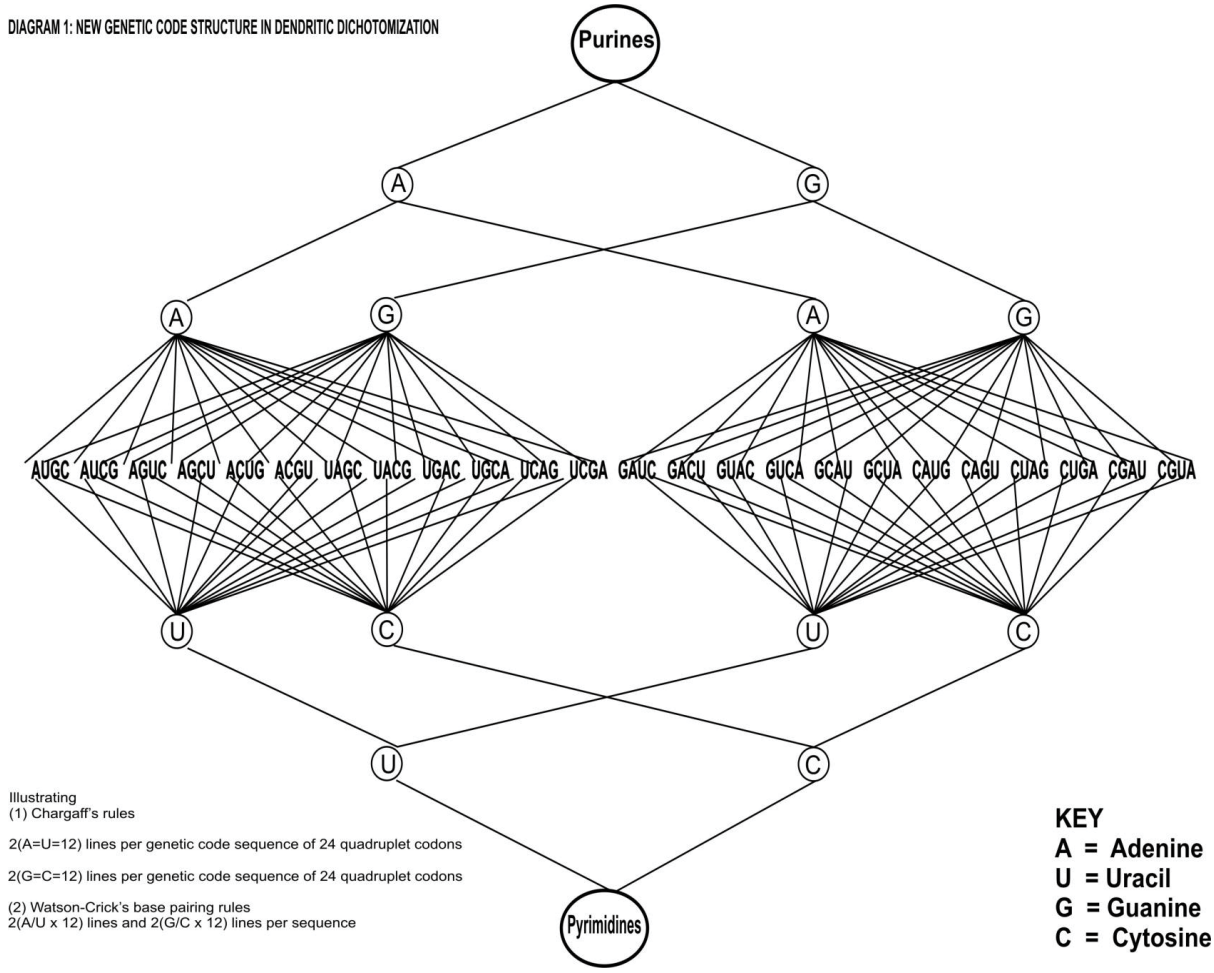


Diagram 1: New genetic code structure in dendritic dichotomization.
Illustrating (1) Chargaff's Rules
2 (A=U=12) lines per genetic code sequence of 24 quadruplet codons
2 (G=C=12) lines per genetic code sequence of 24 quadruplet codons (2) Watson-Crick's base pairing rules
2 (A/U × 12) lines and 2(G/C × 12) lines per genetic code sequence 24 quadruplet codons

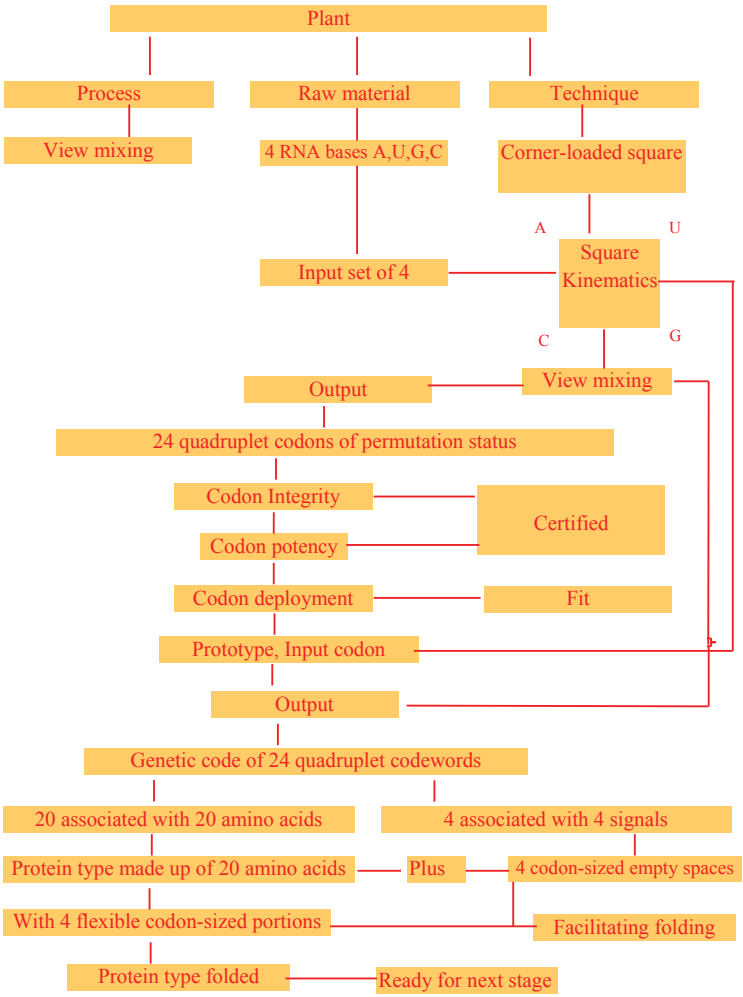


Figure 5: Combinatorial essentials of theoretical plant flow-chart for protein synthesis and primary folding of protein type in box-graphics.

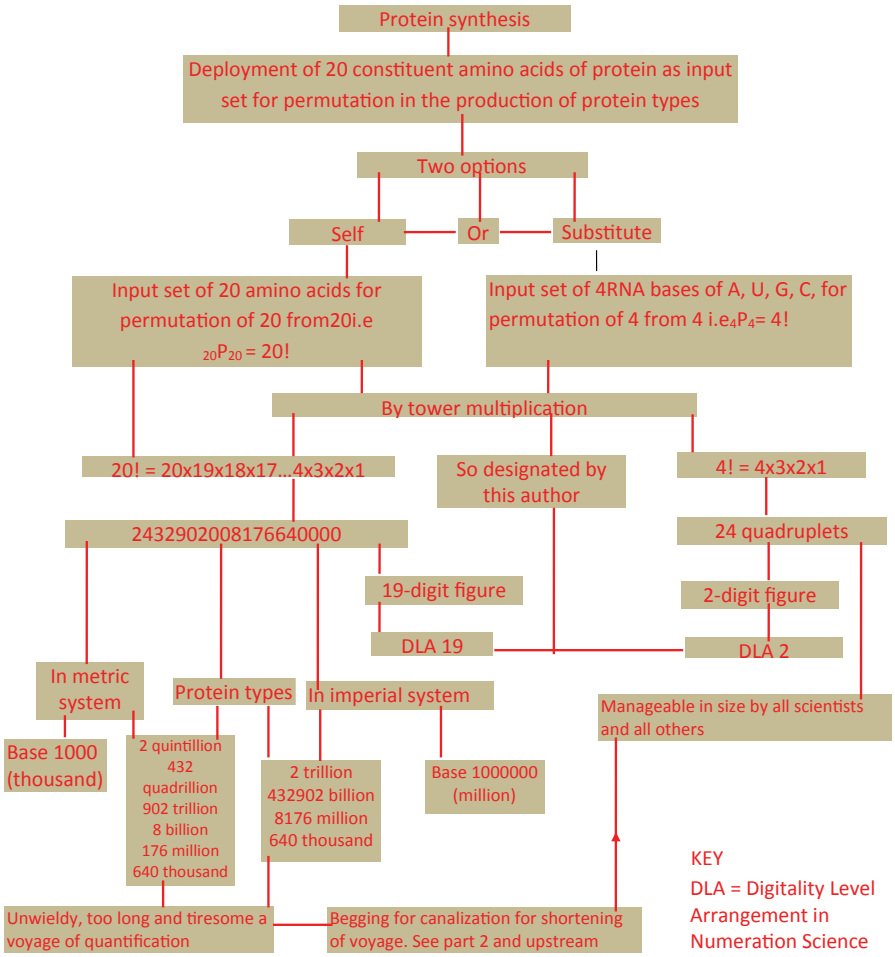


Figure 6a: One of two options for input set in box-graphics protein synthesis, why the genetic code? Protein synthesis: one of two options concerning deployment of 20 constituent amino acids of protein as input set for permutation for combinatorial alignment in the production of protein types; self or substitute:

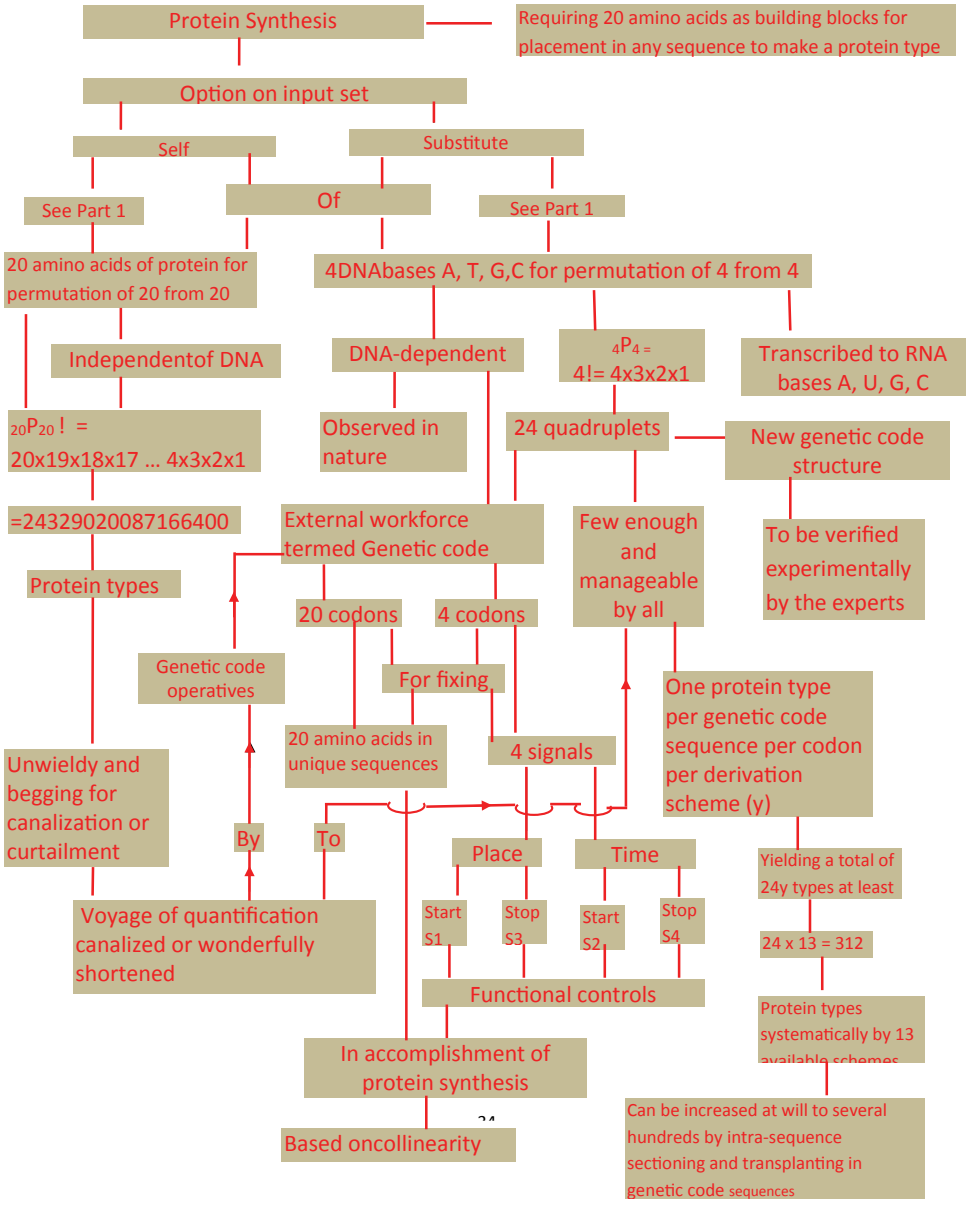


Figure 6b: Protein synthesis, the substitute option of 4 DNA bases A, T, G, C as input set for permutation in box-graphics.

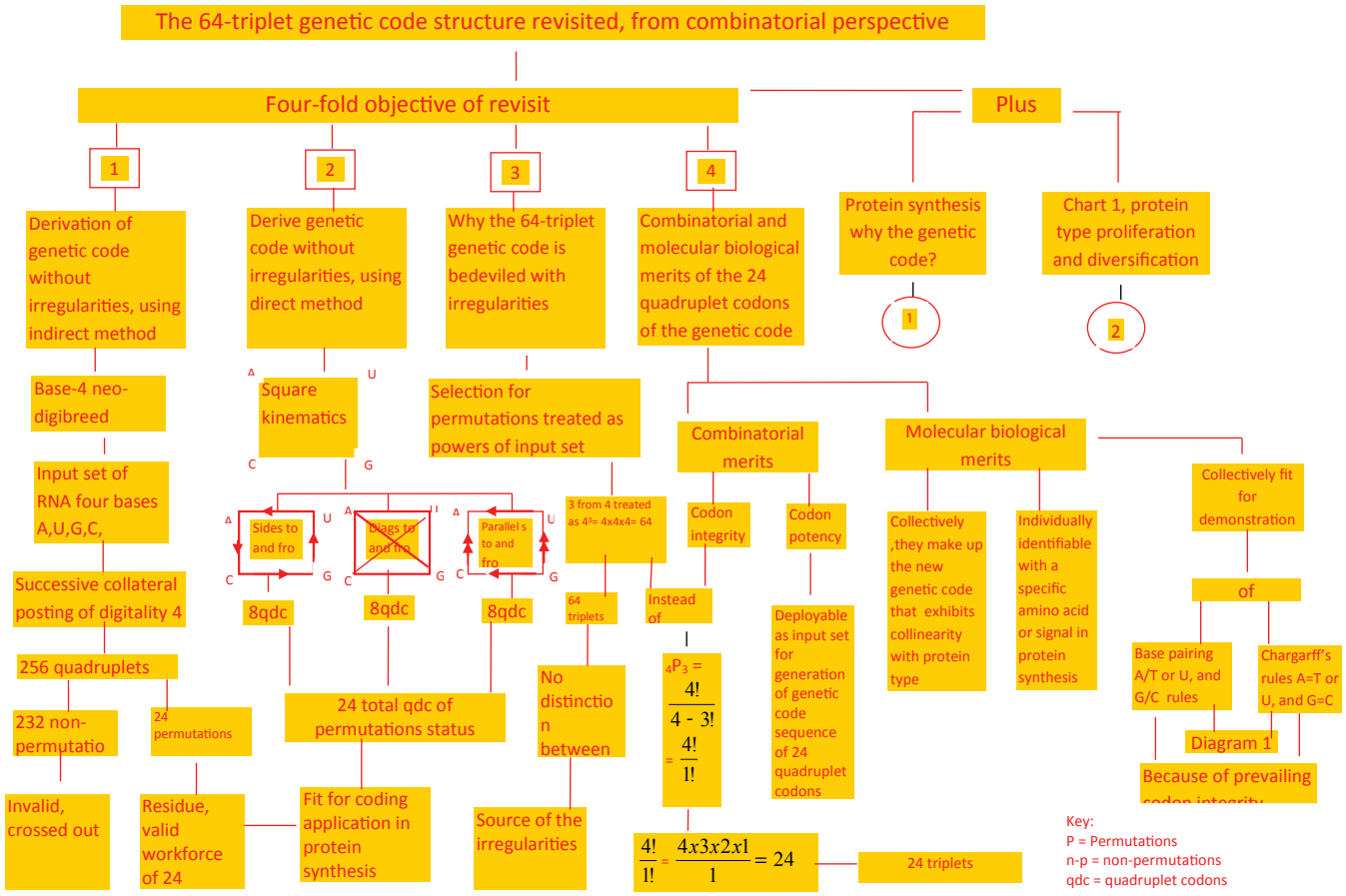


Figure 7a: The 64-triplet genetic code structure revisited, from combinatorial perspective: Graphical abstract part 1.

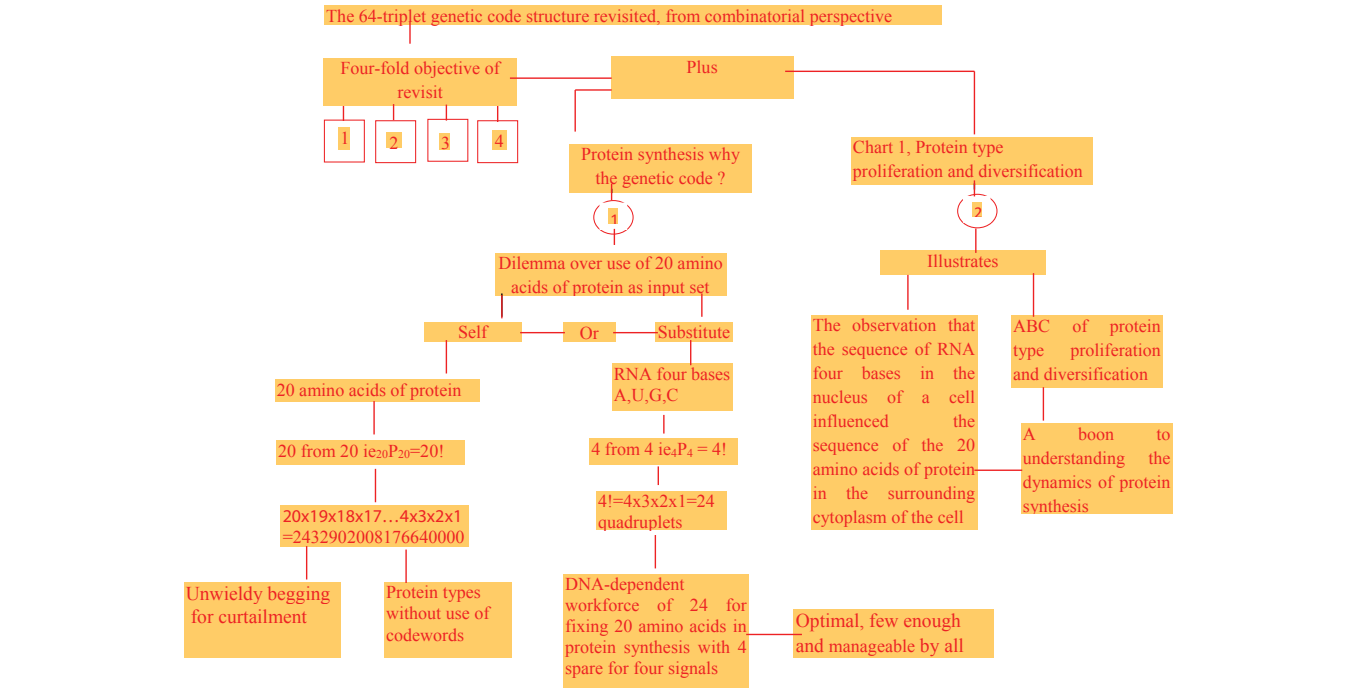


Figure 7b: The 64-triplet genetic code structure revisited, from combinatorial perspective: Graphical abstract part 2.

Conclusion

Collinearity evident in the one-to-one correlation between the 24 quadruplet code words and the 20 amino acids of protein and 4 control signals in protein synthesis in plants and animals as portrayed in Chart 3 is the sure evidence of the combinatorial correctness or accuracy of this new genetic code combinatorial structure yielded by this revisit of the degenerate 64-triplet genetic code structure Figures 7a and 7b. This new genetic code version is wholeheartedly recommended to experimental experts in molecular biology etc. for spelling in order to render it fit for coding application in protein synthesis studies.

Highlights

- Let it be noted at this point that the 64-triplet genetic code structure begging for combinatorial revisit is immature and a mixture of 40 non-permutation and 24 permutation triplets which are as incompatible as weeds and crops during harvest.
- The genetic code structure of 24 permutation quadruplets is the combinatorial answer to the raising of enough code words from the four-letter alphabet of the RNA four bases A, U, G, C. (Adenine, Uracil, Guanine and Cytosine).
- The diversification of protein types follows from the variation of sequence of 24 output factorial complements being the code words as influenced by changes of sequence of bases in the input set of four.
- With a quadruplet input set of RNA bases whether in the indirect method of base-4 neo-digibreed scheme or direct method of *Square Kinematics*, the combinatorially correct output factorial complements of 4 from 4 can only be 24 permutation quadruplets given by factorial $4=4!=4 \times 3 \times 2 \times 1=24$ quadruplets.
- The new genetic code structure of 24 quadruplets affords collinearity, one-to-one correspondence, with the 20 amino acids of protein/4 control signals for effective protein synthesis characterized by protein type diversification.

References

1. Becker WM (1986) The world of the cell. Glossary, The Benjamin/Cummings Publishing Company, Inc. 2727 Sandfill Road, Menlo Park, California 94025, p 529.
2. Jill W (1988) Prentice hall life science, Prentice Hall, A Division of Simon & Schuster Englewood Cliffs New Jersey 07632.
3. Wayne MB (1986) The world of the cell. Glossary p. 529. The Benjamin/ Cummings Publishing Company, Inc.2727 Sandfill Road, Menlo Park, California 94025, Glossary p: 853.
4. Ibid Glossary p 819.
5. Ibid Glossary p 93.