Acta Scientific MEDICAL SCIENCES (ISSN: 2582-0931)

Volume 7 Issue 12 December 2023

Research Article

Quadruplet Expanded DNA (QED) Genetic Code for Eukaryotic Cells

Rama Shankar Singh*

GEN-MED R & D, LLC, Orlando, FL 32817, USA

***Corresponding Author:** Rama Shankar Singh, GEN-MED R & D, LLC, Orlando, FL 32817, USA. email: gen-med@comcast.net

DOI: 10.31080/ASMS.2023.07.1720

Received: October 16, 2023

Published: November 28, 2023

Abstract

Genetic code translates eukaryotic cell genes into proteins for maintaining a homeostatic state. However, gene variants, transcription, and splicing errors yield dysfunctional proteins causing monogenic rare, multigene cancers and neurodegenerative diseases. The triplet genetic code encodes a protein but lacks gene, transcription, and splicing controls. Furthermore, alternative orthogonally expanded genetic codes failed to synthesize proteins using canonical amino acids. The QED codon was developed to overcome these limitations. While verifying the triplet genetic code, 1968 Medicine Nobel laureate H.G. Khorana observed that self-complementarity forming adjacent bases, Poly r-AU, did not promote polypeptide formation, a noncoding trait. The QED noncoding codons have similar traits. Here, the QED codon model is assumed to comprise all four DNA bases (T, C, A, and G); the code is position-independent and symmetric. The self-complementarity forming adjacent bases (AU) and (C G) with any two NN (N any T, C, A, and G) bases are noncoding. Under these QED assumptions, 256 quadruplets fall into two groups: 20 independent protein-encoding codons and 35 independent noncoding codons applicable to regulating and controlling synthesis, transcription, and splicing processes. Steps to correct dysfunctional proteins are described, anticipating strategies for developing cures for monogenic rare, multigene cancers and neurodegenerative diseases.

Keywords: Eukaryote; Quadruplet; Expanded; Genetic Coding; Nondegenerate; Prokaryote; Viruses

Introduction

Gene encoding is the most critical step in translating mRNA gene information into proteins at the ribosome to maintain the homeostatic state of cells. Genetic code development occurred in two distinct periods: pre-1970 and post-1970. The pre-70 triplet genetic code had protein-encoding and limited START and STOP control. Post-70 split gene discovery into eukaryotic cells required transcription, splicing, protein encoding, noncoding regulation, and control. The ribosome's structure became equally critical in decoding and protein synthesis. Expanded orthogonal codons were also developed using unnatural amino acids. Here, the innovative quadruplet expanded DNA (QED) genetic code for eukaryotes is presented. The QED genetic codon tables for protein-coding and noncoding regulatory and controls are generated. Steps are described for curing rare monogenic diseases and cancers by correcting dysfunctional proteins causing these diseases.

The pre-1970 triplet genetic code applies to prokaryotes and viruses.

The birth of triplet coding is closely related to the DNA structure [1,2]. Crick., *et al.* [3] established the DNA structure with four (T, A, C, and G) bases: T: A bases and C: G bases, naturally forming complementarity pairs. Crick introduced the central dogma of biology. The DNA is considered hereditary material. Protein

synthesis occurs from DNA to mRNA to protein, and one gene-one protein. The codons translate mRNA genetic information into proteins.

In 1961, Crick described further the genetic code and the proteins [4]. In 1963 [5,6], he proposed a triplet genetic code. In 1968, Nobel Prize in Medicine and Physiology [7] was awarded to Robert W. Holley, Har Gobind Khorana, and Marshal W. Nirenberg for verifying the 64 triplet codons: 61 triplet protein encoding, 2 STOP, and one START codon. Khorana used the synthesis process [8], while Nirenberg used the enzymatic binding process [9,10]. Holley [11] established a tRNA structure with attached amino acids and anticodons. At the ribosome, tRNA anticodons form Watson-Crick (WC) pairs with the first two mRNA triplet bases and a wobbly pair with the third base causing degeneracy in the coding.

The triplet encoding has no gene control. Jacob and Monod [12] developed operons to control gene and enzyme synthesis via operator, promotor, regulator, and repressor. The regulator gene initiates the process. In transcription, the regulator and repressor set the operator on or off for RNA polymerization. Two widely used operons are lac and trp. The lac operon has a negative default control and is demonstrated by the lactose digestive process. When lactose is absent, no action occurs, but when present, the lac operon controls the gene to synthesize the enzymes to digest the lactose.

The trp has a positive default control. When tryptophan is present, nothing happens, but Trp controls the synthesis of tryptophan when absent. The 1965 Nobel Prize in Physiology or Medicine [13] was awarded to François Jacob, André Lwoff, and Jacques Monod for their developing genetic control of enzyme and virus synthesis.

Shortcomings of the pre-70 triplet genetic code

The triplet genetic code is nonoptimal, invokes the wobble hypothesis, and lacks eukaryote control. Additionally, since the number of corresponding tRNAs is insufficient for twenty amino acids, iso-tRNA decode multiple amino acids.

The Viruses violated the first rule of central dogma of biology by starting virus mRNA, then RT, cDNA, mRNA, and protein. The split gene and alternate splicing broke the second one gene–one protein rule by synthesizing one gene -multiple proteins.

Post-1970 DNA code for eukaryotes – splicing and transcription

Eukaryote Splicing

In approximately 1977, it became known [14,15] that less than 2% of DNA bases encode proteins, and the remaining is noncoding or junk. Genes were not continuously distributed but coding portions (exons) separated by noncoding parts (introns). The splicing process separates exons from introns. The 1993 Nobel Prize in Physiology or Medicine [16] was awarded to Richard J. Roberts and Phillip A. Sharp for discovering "split genes." Several unique proteins can be synthesized using alternate splicing from a single gene [17,18]. Thus, Eukaryotic cells require transcription, splicing, and various regulatory and control processes, including epigenetics.

Eukaryote transcription

Roger Kornberg elucidated the detailed transcription process using Baker's yeast as a eukaryotic model. The transcription process starts with the TATA box and several transcription factor-binding proteins, mediators, promotors, activators, and other controlling elements. DNA transcription (RNA polymerization) yields Pol-I rRNA, Pol-II mRNA, and Pol-III tRNA. Ribosomes and tRNAs are synthesized using Pol-I rRNA and Pol-III tRNA, respectively. At the ribosome, Pol-II mRNA codons are translated into protein. Kornberg [19] was awarded the 2006 Chemistry Nobel Prize for his "fundamental studies of the molecular basis of eukaryotic transcription".

Splicing and Transcription errors

Transcription errors cause several human diseases [20]. Splicing errors also cause diseases [21,22].

The ribosome protein-making factory and the gene decoding

In the post-70 era, the ribosome structure was critical to understanding protein synthesis and decoding. In 1955, Palade [23] first identified this organelle, later named ribosome. Venkatraman Ramakrishnan, Thomas A. Steitz, and Ada Yonath determined the detailed ribosome structure by x-ray. The 2009 Nobel Prize in Chemistry was awarded to them "for studies of the structure and function of the ribosome" [24]. Eukaryotes, prokaryotes, and archaea have similar structures but differ in sizes and ribose protein ratios. The protein synthesis at the ribosome

was illustrated using the ribosome structure [25-27].

Later, V. Ramakrishnan described the race to decipher the secret of ribosomes in his book "Gene Machine" [28]. The ribosomal decoding of the codon at the third wobble position [29,30] are flexible enough to accommodate a codon at the fourth position. The ribosome's decoding, translocation, and extension activities ensure the proper synthesis of a protein.

Ribosome structure is equally critical in controlling bacterial-antibiotic interactions [31].

In the post-1970 era expanded codons: orthogonally expanded quadruplet [32-35], sextuplet [36], and octuplet [37] genetic codes were developed using unnatural amino acids.

The orthogonal expanded codons have yet to synthesize proteins using canonical amino acids.

Methods

Model: Quadruplet expanded DNA (QED) genetic code for eukaryotes

The QED genetic code is developed by reviewing H.G. Khorana's triplet code verification [7].

- Khorana observed [8] that self-complementary forming adjacent bases AU, poly-rAU, and CG, poly-rCG, are noncoding.
- Poly r-trinucleotides yielded gene encoding and noncoding controls STOP (UAA, UAG, and UGA) and START (AUG) codons.
- The G base position in two STOP UGA and UAG codons seems to be position independent and symmetric, i.e., U(GA): U(AG), with no sensitivity at the third base.

The QED genetic code for eukaryote cells has the following assumptions:

- All four DNA (A, T, C, and G) bases are involved; in mRNA, T is replaced by U.
- Base positions are independent; i.e., for any A and B, AB and BA are equivalent.
- Base positions are symmetric; i.e., for any A and B, (AB) and (BA) are synonymous.
- Self-complementarity forming adjacent base pairs with any two adjacent NN (N any A, T, C, or G) bases, (AT) NN and (CG) NN, is noncoding, controlling and regulating the processes. Following assumption (3), (AT)(NN)) and (NN)(AT)) are synonymous; likewise, (CG)(NN) and ((NN)(CG)) are synonymous.

The detailed methods for generating QED eukaryote codons

Under assumptions (1) to (3), codons are arranged in a square symmetric matrix. Any N x N square symmetric matrices have N x (N+1)/2 independent elements, and element M (I, J) is synonymous with M (J, I), where I and J are the rows and columns of the matrix, respectively.

Four DNA bases arranged in a 4 x 4 square symmetric matrix yield 4 x (4+1)/2 = 10 independent elements, which set in a 10 x 10 square symmetric matrix yield 10 x (10+1)/2 = 55 independent elements. Under the 4th QED assumption, these fifty-five elements result in 20 independent protein-encoding elements and thirty-five independent noncoding elements for gene regulation, and control.

	T	C	A	G
T	TT	(TC)	(TA)	(TG)
C		CC	(CA)	(CG)
A			AA	(AG)
G				GG

Table 1: (a) Four DNA (T, C, A, and G) bases arranged in a 4x4 square symmetric matrix.

Only the upper 10 symmetric independent elements of matrix M (I, J) are shown. The lower elements M (J, I) are synonymous to M (I, J) = M (J, I), where row I = 1,2,3 and 4, and column J = 1,2,3 and 4. Additionally, elements M (I, J) and M (J, I) are synonymous. Thus, (TC):(CT), (TG):(GT), (CA): (AC),(AG):(GA); (TA):(AT) and (CG):(GC) in(red) are synonymous with each other. Applying the 4th QED codon assumption, the 8 elements are part of coding, and the two elements (TA) and (CG) in (red) are the part of noncoding.

Next, the 10 symmetric and independent elements of Table 1(a) are arranged in Table 1(b). The coding elements are in, the noncoding elements in (red) (Table 1b).

	TT	CC	AA	GG	(CT)	(AC)	(TG)	(AG)	(TA)	(CG)
TT	TTTT	(TT)(CC)	(TT)(AA)	(TT)(GG)	TT(CT)	TT(AC)	TT(TG)	TT(AG)	TT(TA)	TT(CG)
CC		CCCC	(CC)(AA)	(CC)(GG)	CC(CT)	CC(AC)	CC(TG)	CC(AG)	CC(TA)	CC(CG)
AA			AAAA	(AA)(GG)	AA(CT)	AA(AC)	AA(TG)	AA(AG)	AA(TA)	AA(CG)
GG				GGGG	GG(CT)	GG(AC)	GG(TG)	GG(AG)	GG(TA)	GG(CG)
(CT)					(CT)(CT)	(CT)(AC)	(CT)(TG)	(CT)(AG)	(CT)(TA)	(CT)(CG)
(AC)						(AC)(AC)	(AC)(TG)	(AC)(AG)	(AC)((TA)	(AC)(CG)
(TG)							(TG)(TG)	(GT)(AG)	(GT)(TA)	(GT)(CG)
(AG)								(AG)(AG)	(AG)(TA)	(AG)(CG)
(TA)	(TA)TT	(TA)CC	(TA)AA	(TA)GG	(TA)(CT)	(TA)(AC)	(TA)(GT)	(TA)(AG)	(TA)(TA)	(TA)(CG)
(CG)	(CG)TT	(CG)CC	(CG)AA	(CG)(GG)	(CG(CT)	(CG)(AC)	(CG)(TG)	(CG)(AG)	(CG)(TA)	(CG)(CG)

Table 1(b): Ten symmetric and independent elements of Table 1(a) set in a 10x10 square symmetric matrix.

Only the upper half of the symmetric and independent coding (bold) and noncoding (red) elements of square matrix M (I, J) are shown. Under 4th QED assumption, any combinations of (AT) NN and (CG) NN (where N is any A, T, C, or G) in (red) are noncoding. The lower half of symmetric matrix M (J,I) can be generated using M(J,I) = M(I,J) (where I = 1,2,3…10, and J = 1,2,3..10). The synonymous codon can be generated using these elements, as illustrated in rows 9 and 10 for columns 9 and 10, respectively.

The twenty bold independent protein-encoding codons from Table 1(b) (replacing T with U for mRNA) and the corresponding synonymous codons are shown in Table 2 (a). In Table 2 (b), the thirty-five, independent noncoding codons (retaining DNA bases) are shown in (red) font.

	Quadrupleu expanded DNA (QED) Codons				
	Coding Codons	Synonymous coding codons, (T>U)			H. B.
1	UUUU	UUUU			8
2	CCCC	CCCC			12
3	AAAA	AAAA			8
4	GGGG	GGGG			12
5	(AA)(CC)	(CC)(AA)			10
6	(UC)CC	(CU)CC	CC(UC)	CC(CU)	11
7	(UG)UU	(GU)UU	UU(UG)	UU(GU)	9
8	(UG)GG	(GU)GG	GG(UG)	GG(GU)	11
9	(CA)CC	(AC)CC	CC(CA)	CC(AC)	11
10	(UU)(GG)	(GG) (UU)			10
11	(AC)(CA)	(AC)(AC)	(CA)(CA)	(CA)(AC)	10
12	(GA)(GA)	(GA) (AG)	(AG)(GA)	(AG)(AG)	10
13	(GU)(GU)	(GU) (UG)	(UG)(UG)	(UG)(GU)	10
14	(GA)GG	GG(GA)	GG(AG)	(AG)GG	11
15	(CA)AA	(AC)AA	AA(CA)	AA(AC)	9
16	UU(UC)	UU(CU)	(UC)UU	(CU)UU	9
17	(AG)AA	AA(GA)	AA(AG)	(GA)AA	9
18	(AA)(GG)	(GG) (AA)			10
19	(CU)(CU)	(CU) (UC)	(UC)(UC)	(UC)(CU)	10
20	(UU)(CC)	(CC) (UU)			10

Table 2a: (a) The twenty protein-coding QED codons and their synonymous codons. For protein synthesis, T in Table 1(b) has been replaced by U for mRNA, Number of Hydrogen Bond (H.B.).

	Noncoding codons	Noncoding Synonymous codons			H.B.
1	(TA)(TA)	(TA)(AT)	(AT)(TA)	(AT) (AT)	8
2	(CG)(CG)	(CG)(GC)	(GC)(CG)	(GC) (GC)	12
3	(AU)GG	GG(AU)	GG(UA)	(UA)GG	10
4	(UG)(AC)	(AC)(UG)	(UG) (CA)	(AC) (GU)	10
5	(UG)(AG)	(GU)(AG)	(UG) (GA)	(GU) (AG)	10
6	(UG)AA	AA(UG)	(GU)AA	AA(GU)	9
7	(UA)(GU)	(GU)(UA)	(UA) (UG)	(GU) (AU)	9
8	(UA)(GA)	(AG)(UA)	(UA) (AG)	(GA) (AU)	9
9	(UA)(GC)	(UA)(CG)	(CG)(UA)	(CG) (AU)	10
10	(UA)AA	AA(UA)	(AU)AA	AA(AU)	8
11	(UA)(AC)	(AC)(UA)	(UA) (CA)	(AC) (AU)	9
12	(TT)(AA)	(AA)(TT)			8
13	(CC)(GG)	(GG)(CC)			12
14	TT(TA)	(TA)TT	(AT)TT	TT(AT)	8
15	TT(AC)	(AC)TT	(CA)TT	TT(CA)	9
16	TT(AG)	(GA)TT	(AG)TT	TT(GA)	9
17	TT(CG)	(CG)TT	TT(GC)	(GC)TT	10
18	CC(TA)	(TA)CC	(AT)CC	CC(AT)	10
19	CC(TG)	(TG)CC	(GT)CC	CC(GT)	11
20	CC(AG)	(AG)CC	(GA)CC	CC(GA)	11
21	CC(CG)	(CG)CC	(GC)CC	CC(GC)	12
22	AA(CT)	(CT)AA	(TC)AA	AA(TC)	9
23	AA(CG)	(GC)AA	(CG)AA	AA(GC)	10
24	GG(CT)	(CT)GG	(TC)GG	GG(TC)	11
25	GG(CG)	(CG)GG	(GC)GG	GG(GC)	12
26	GG(AC)	(AC)GG	(CA)GG	GG(CA)	11
27	(AC)(CG)	(CA)(CG)	(CA)(GC)	(AC) (GC)	11
28	(AC)(AG)	(AC)(GA)	(CA)(GA)	(CA) (AG)	10
29	(AG)(CG)	(GA)(CG)	(AG)(GC)	(GA) (GC)	11
30	(CT)(TA)	(TC)(TA)	(CT)(AT)	(TC) (AT)	9
31	(CT)(CG)	(TC)(CG)	(CT)(GC)	(TC) (GC)	11
32	(CT)(AC)	(TC)(AC)	(CT)(CA)	(TC) (CA)	10
33	(CT)(AG)	(TC)(AG)	(CT)(GA)	(TC) (GA)	10
34	(CT)(TG)	(TC)(TG)	(CT)(GT)	(TC) (GT)	10
35	(GT)(CG)	(TG)(CG)	(GT)(GC)	(TG) (GC)	11

Table 2(b): Thirty-five QED noncoding synonymous codons from Table 1 (b).

From Table 2(a) and 2(b), the numbers of hydrogen bonds forming in coding and noncoding codons are respectively shown in Figure 1 (a) and (b).

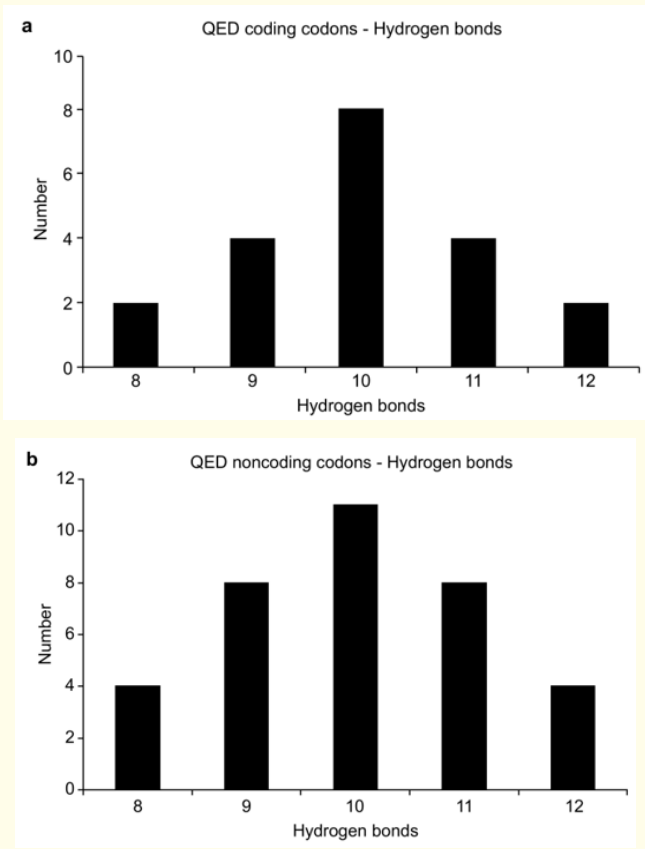

Figure 1

QED protein-encoding codon assignments

The QED protein-encoding codons encode proteins in both eukaryotes and prokaryotes as protein synthesis process is similar. Therefore, the tentative QED protein-encoding codon assignment could use the already verified triplet coding using its first two bases, ignoring the degeneracy and third base wobble. Therefore, the triplet codon table was rearranged with amino acids, degenerate codons, and corresponding tRNAs by imposing the 4th QED codon code assumptions in Table 3.

Amino acid	Triplet mRNA codons under QED constraint and tRNA anticodons		
	Triplet codon/QED	Compressed form	tRNA-anticodon (38,39)
Ala/A	~~GCU, GCC, GCA, GCG~~	~~GCN~~, GCA?	UGC
Arg/R	~~CGU, CGC, CGA, CGG,~~ AGA, AGG	AGR	CCG, ACG
Asn/N	~~AAU~~, AAC	AAC	GUU
Asp/D	~~GAU, GAC~~	~~GAY~~, GAC?	GUC
Cys/C	UGU, ~~UGC~~	UGU	GCA
Gln/Q	CAA, ~~CAG~~	CAA	UUG
Glu/E	GAA, GAG	GAR	YUC
Gly/G	GGU, ~~GGC~~, GGA, GGG	GGD	NCC
His/H	~~CAU~~, CAC	CAC	GUG
Ile/I	~~AUU, AUC, AUA~~	~~AUH~~, AUC?	GAU
Leu/L	~~UUA~~, UUG, CUU, CUC, ~~CUA, CUG~~	UUG, CUY	YAA
Lys/K	AAA, AAG	AAR	YUU
Met/M	~~AUG*~~	~~D~~, AUG?	CAU
Phe/F	UUU, UUC	UUY	RAA
Pro/P	CCU, CCC, CCA, ~~CCG~~	CCH	KGG
Ser/S	UCU, UCC, ~~UCA, UCG, AGU, AGC~~	UCY	GGA
Thr/T	~~ACU~~, ACC, ACA, ~~ACG~~	ACM	NGU
Trp/W	UGG	UGG	CCA
Tyr/Y	~~UAU, UAC~~	~~UAY~~, UAC?	GUA
Val/V	GUU, ~~GUC~~, G~~UA~~, GUG	GUK	NAC
START	AUG	AUG	
STOP	UAA, UAG, UGA	UAR, UGA	

Table 3: Amino acids, triplet mRNA codons and tRNA anticodons, and stricken out disallowed triplet codons under the 4th QED codon assumptions.

N: Any U, C, A or G; Purine: R = A or G; Pyrimidine: Y = T (U) or C;?: matching tRNA

D: not C; H: not G; K: G or U; M: A or C

QED protein-coding codons are assigned using Tables 2 (a) and 3.

In Table 3, Nirenberg showed [9,10] that polyU, polyA and polyC encode the amino acids Phe, Lys and Pro, respectively. This established a direct link among mRNAs, tRNAs, amino acids, codons and anticodons in protein synthesis at ribosomes. Additionally, in [9,10] oligo chain lengths of 3 and 4: $(oU)_3$ and $(oU)_4$ showed nearly the same activities. Therefore, it is reasonable to assume that if triplet UUU can encode Phe, quadruplet UUUU could also encode Phe. Following this reasoning, LLLL-Lys and CCCC-Pro have been assigned. Since GGG in Table 3 encodes Gly, GGGG-Gly has also been assigned. Thus, four QED codons have been assigned as follows: QED: UUUU – Phe; AAAA –Lys; CCCC- Pro; and GGGG-Gly are listed in Table 4 (a).

Next, sixteen QED codons are assigned following the Table 3 triplet codon assignments. In Crick's original proposal, codons of only two bases could encode only sixteen amino acids. Hence, he added a third base, creating codon degeneracy and allowing the third base to form a dangling bond with the first base of the tRNA anticodon. For QED codon assignments, the first two bases of the triplet codon of each amino acid in Table 3 are compared with the first two bases of the QED protein-coding codons in Table 2(a). When a match occurs, the matching QED codon is assigned to that amino acid. Following this method, the QED codons are assigned as follows: Table 3, Arg/R–AGA, AGG: In this case, if G is added to AGA and A is added to AGGA, then under QED assumptions 2 and 3, (AG)(GA) will represent both. In Table 2 (a), element # 12 (AG)(GA) matches this outcome. Thus, in Table 4(a), QED (AG)(GA)-Arg/R is assigned.

Table 3, Asn/N-AAC: Under QED 4th coding assumption, only C can be added at the fourth position, resulting in AA (CC). Element #5 of Table 2 (a) matches this outcome. Thus, in Table 4 (a), AA (CC)–Asn/N is assigned.

Table 3, Cys/C-UGU: Under the QED coding constraint, only U can be added, resulting in UGUU. Element #7 of Table 2(a) matches this outcome. Thus, in Table 4 (a), (UG) UU-Cys/C is assigned.

Table 3, Gln/Q-CAA: Under the QED rules, U and G are not allowed. Only A can be added, resulting in (CA) AA. Element #15 of Table 2 (a) matches this outcome and (CA) AA-Gln/Q is assigned in Table 4 (a).

Table 3, Glu/E–GAA, GAG: Here, either A or G can be added to either codon, but adding A to GAA will result in a lower preferred bonding energy. Thus, GAAA is preferred. Isoform element #17 of Table 2 (a) matches this outcome and is assigned (GA) AA-Gln/Q in Table 4 (a).

Table 3**,** His/H-CAC: under the QED rules, only C can be added in the fourth position, resulting in CACC. Element # 9 of Table 2 (a), (CA) CC matches this outcome and is assigned (CA) CC–His/H in Table 4 (a).

Table 3, Leu/L-UUG, CUU, and CUC: here at the third position, there are one purine and two pyrimidines. Thus, a pyrimidine (U or C) will be preferred. Since U will require a lower bonding energy than C, U is selected for the fourth position, leading to (CU)(CU). In Table 2 (a), element # 19, (CU)(CU) matches this and is assigned (CU)(CU)-Leu/L in Table 4 (a).

Table 3, Ser/S-UCU, UCC: as in the previous case, either U or C can be added at the fourth potion. Adding U to UCU will result in a lower energy, (UC) UU. Element # 16 of Table 2 (a) matches this outcome and is assigned (UC) UU-Ser/S in Table 4 (a).

Table 3, Thr/T-ACC, ACA: Following the previous reasoning, A is added to ACC and C to ACA, transforming these two codons in to the same codon (AC)(CA). Element # 11 of Table 2 (a) matches this outcome. Therefore, (AC)(CA)-Thr/T is assigned in Table 4 (a).

Table 3, Trp/W-UGG: It is safe to just add G at the fourth position, resulting in UGGG. Element # 8 of Table 2 (a), (UG) GG matches this outcome and is assigned as (UG) GG–Trp/W in Table 4 (a).

Table 3, Val/V-GUU, GUG: As in the two previous cases, G is added to GUU, and U is added to GUG, resulting in the same codon (GU)(UG). Element # 13 of Table 2 (a) matches this and is assigned as (GU)(UG) – Val/V in Table 4 (a).

Table 4 (b), summarizes the QED protein encoding codon assignment of Table 4 (a) with number of hydrogen bonds as the new QED protein encoding codon table.

QED codons encoding amino acids in Table 4 (b) have an exciting feature.

Amino Acids	mRNA under QED	QED codons	Ref./ comm.
Arg/R	AGA, AGG	(AG)AA	38
Asn/N	AAC	(AA)(CC)	38
Cys/C	UGU	(UG)UU	38
Gln/Q	CAA	(CA)AA	38
Glu/E	GAA, GAG	(GA)(GA)	38
Gly/G	GGU, GGA, GGG	GGGG	9,10
His/H	CAC	(CA)CC	38
Leu/L	UUG, CUU, CUC	(CU)(CU)	38
Lys/K	AAA, AAG	AAAA	9,10
Phe/F	UUU, UUC	UUUU	9,10
Pro/P	CCU, CCC, CCA	CCCC	9,10
Ser/S	UCU,UCC	(UC)UU	38
Thr/T	ACC, ACA	(AC)(CA)	38
Trp/W	UGG	(UG)GG	38
Val/V	GUU, GUG	(GU)(GU)	38
Ala/A	~~GCN~~?	(GG)(AA)**	
Asp/D	~~GAY~~?	(GA)(GG)**	
Ile/I	~~AUH~~?	UU(GG)**	
Met/M	~~AUG~~?	(UC)CC**	
Tyr/Y	~~UAY~~?	(UU)(CC)**	
START	AUG	Noncoding	Regulatory
STOP	UAA, UAG, UGA	Noncoding	Regulatory

Table 4(a): The QED encoding codon assignments.

**To be assigned; (?): to be determined.

Amino acids	QED-Codons	H.B.	QED-Codons	Amino acids
Phe	UUUU	8	AAAA	Lys
Cys	(UG)UU	9	(CA)AA	Gln
Ser	(UC)UU	9	(AG)AA	Arg
Thr	(AC)(CA)	10	(GU)(GU)	Val
Asn	(AA)(CC)	10	(UU)(GG)	* Met
**	(UU)(CC)	10	(GG)(AA)	**
Glu	(GA)(GA)	10	(CU)(CU)	Leu
Trp	(UG)GG	11	(CA)CC	His
**	(UC)CC	11	(GA)GG	**
Pro	CCCC	12	GGGG	Gly

Table 4(b): The QED protein encoding codon table arranged in H.B. ascending order.

*Met, to be verified, ** to be assigned: Ala/A, Asp/D, lle/l, and Tyr/Y.

The anticodon of the QED codon encoding an amino acid is the encoding QED codon of the other amino acid. For example, UUUU encodes Phe, and its anticodon AAAA encodes Lys. (UG)UU encodes Cys, and its anticodon is (AC)AA which is synonymous with (CA)AA, see Table 2 (a) number 9. The same trait is valid for the remaining codons.

Based on the QED codon-anticodon relation, a possibility exists that only ten tRNA may be needed to synthesize proteins using canonical amino acids.

Multiple triplet codons code the same amino acid, but one tRNA decodes many amino acids. However, AUG encodes both control SART and amino acid Met. What makes this dual role? Also, Met is not found first amino acid in every protein. What is the mechanism of Met, where the first amino acid but clipped?

It has been reported [40] that triplet GUG and UUG encode Met. Thus following the prior procedure, if U is added to GUG, and G to UUG, then the QED codon (UU) (GG) will cover both codons. Element # 10 of Table 2(a) matches the outcome, and tentatively *(UU) (GG) – Met is assigned. Since AUG has been assigned a noncoding START codon in QED, this double role dilemma will not arise.

Noncoding QED codon assignment leading to regulation and control

The thirty-five noncoding QED codons from Table 2 (b) are anticipated to regulate protein synthesis (as in triplet coding), and eukaryotic transcription, and splicing processes. Following the protein-coding assignment procedure of QED codons described above, the verified triplet START and STOP codons are used to assign the corresponding QED codons. Since the information in Table 2 (b) is provided in DNA bases, T has been replaced by U in the QED START and STOP codons. The assigned noncoding QED codons are listed in Table 5.

Table 2 (b), first element: (TA)(TA).

In eukaryotes, transcription and splicing are the critical pre-mRNA processing steps to produce rRNA, tRNA, and mRNA for protein synthesis. Transcription always starts at the TATA box.

Table 2(b) shows element #1 (TA)(TA), and is assigned to initiate the transcription process and listed in Table 5.

Table 2(b), 2nd element: (CG)(CG).

The splicing process separates protein-coding exons from noncoding introns and is unique to eukaryotes. The (CG)(CG) element and G + C-rich bases are used to locate exon–intron interfaces, and splicing then separates them. Furthermore, alternative splicing makes it possible for one gene to encode multiple proteins. Therefore, the (CG)(CG) element of Table 2(b) is assigned for controlling splicing processes and is listed in Table 5.

Table 2 (b), assignments of the following 3 to 11 elements.

In Table 1 (b), among 35 noncoding codons, 10 are (AU) NN, and 10 are (CG) NN (where N is A, T, C, or G) listed in Table 2(b). The remaining fifteen noncoding codons are mixed combinations.

For the QED START and STOP codon assignments, the triplet START and STOP codons of Tables 3 are used as guides. Additionally, the T bases of these nine elements in Table 2(b) have been replaced by U.

START

START–AUG triplet matches the first two bases of the third element in Table 2 (b). Thus, QED START-(AU) GG is assigned and listed in Table 5.

STOP

In Table 3, STOP triplets include three codons: UGA, UAG, and UAA.

QED STOP: The first two bases of elements 4 to 6 of Table 2 (b) match the first two bases of the UGA triplet. Thus, QED STOP-(UG) (AC), -(UG)(AG), and -(UG)(AA) are assigned in Table 5.

Since (UG) AA has lower bonding energy, it is assigned STOP. The other two are assigned as Regulatory or STOP.

The first two bases of elements 7 to 9 of Table 2 (b) match the first two bases of the triplet UAG.

Thus, QED STOP–(UA)(GU), -(UA)(GA), and -(UA)(GC) are assigned in Table 5.

Following the previous procedure, (UA)(GA) is assigned STOP and the other two as Regulatory or STOP.

Table 2 (b), 10th, and 11th elements.

The two bases of the 10th and 11th elements match the first two bases of triplet UAA. Thus, QED STOP-(UA)AA and -(UA)(AC) are assigned in Table 5. Following the previous procedure, (UA) AA is assigned STOP and (UA)(AC) as Regulatory or STOP.

The assignment of the remaining twenty-four G+C- and T+A-rich QED regulatory noncoding codons will require further work.

**	Triplet Codons	Non-coding QED Codons	QED Regulatory and Control		
1	No provision	(TA)(TA)	TATA Box - Transcription start		
2	No provision	(CG) (CG)	(CG)(CG), Exon/Intron Interface		
3	START-AUG	(AU)GG	START	Comments	
5	STOP-UGA (OPAL)	(UG) (AG)	STOP		
8	STOP-UAG(AMBER)	(UA) (GA)	STOP		
10	STOP-UAA(OCHER)	(UA)AA	STOP		
4		(UG) (AC)	Regulatory	*	STOP
6		(UG)AA	Regulatory	*	STOP
9		(UA) (GC)	Regulatory	*	STOP
7		(UA) (GU)	Regulatory	*	STOP
11		(UA) (AC)	Regulatory	*	STOP
12		(TT) (AA)	Regulatory	*	
13		(CC) (GG)	Regulatory	*	
14		TT(TA)	Regulatory	*	
15		TT(AC)	Regulatory	*	
16		TT(AG)	Regulatory	*	
17		TT(CG)	Regulatory	*	
18		CC(TA)	Regulatory	*	
19		CC(TG)	Regulatory	*	
20		CC(AG)	Regulatory	*	
21		CC(CG)	Regulatory	*	
22		AA(CT)	Regulatory	*	
23		AA(CG)	Regulatory	*	
24		GG(CT)	Regulatory	*	
25		GG(CG)	Regulatory	*	
26		GG(AC)	Regulatory	*	
27		(AC) (CG)	Regulatory	*	
28		(AC) (AG)	Regulatory	*	
29		(AG) (CG)	Regulatory	*	
30		(CT) (TA)	Regulatory	*	
31		(CT) (CG)	Regulatory	*	
32		(CT) (AC)	Regulatory	*	
33		(CT) (AG)	Regulatory	*	
34		(GT) (CG)	Regulatory	*	
35		(GT) (AG)	Regulatory	*To be assigned	

Table 5: QED regulatory noncoding codon assignments, ** Table 2 (b), numbers.

Incurable monogenic rare, multigenic cancer, and neurodegenerative diseases and vaccines

Gene variants, transcription, and splicing errors produce dysfunctional proteins causing the disease. More than 7,000 monogenic rare diseases have no cure, only management of symptoms.

A similar situation is observed for multigenic cancers. Once cancer is detected, the treatment is initiated with surgery, followed by radiation and chemotherapy. The goal has been to extend the five-year life.

In rare diseases, dysfunctional protein correction requires the replacement of incorrect amino acids with the correct

ones. However, the degenerate triplet codon makes selecting a unique codon among the degenerate ones a foremost hurdle. The nondegenerate protein-encoding QED codons have no such limitation. In gene therapy, variant genes are first corrected at the DNA level with CRISPR gene editing tools. Normal proteins are generated next to replace dysfunctional proteins.

For cancer cure, no biological technique exists to access cancerous cells selectively, a foremost hurdle that must be overcome. Lack of gene, transcription, and splicing control in triplet code might have prevented the development of such technique for eukaryotes. The eukaryote QED genetic code has the potential to develop such a technique.

Protein synthesis to correct dysfunctional proteins

QED codons translate the genetic information in mRNA into proteins like triplet code does at the ribosome. The translation process is the same in eukaryotes, prokaryotes, and viruses, but the starting and intervening steps differ, as shown in Figure 2a-c. The different roles of the QED codons in control and translation are shown.

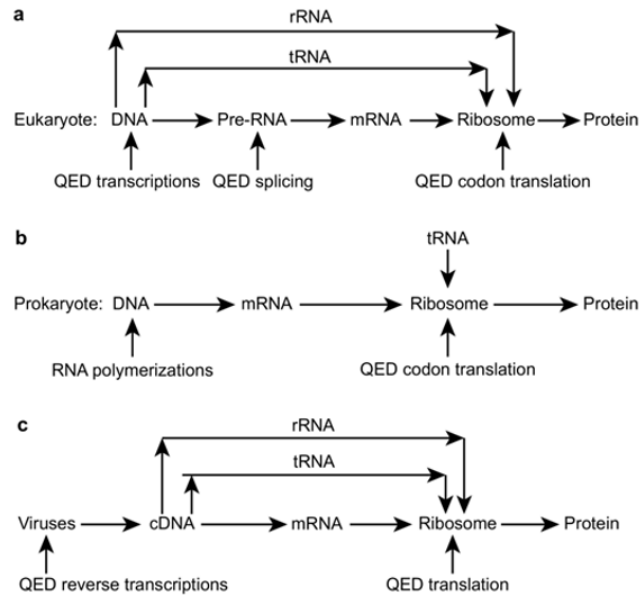

Figure 2: Protein synthesis process.

Dysfunctional proteins causing diseases could be corrected at the protein level or gene therapy at the DNA level. The steps are illustrated in Figure 3 and 4.

(a) The protein production pathway in eukaryotes has additional transcription, splicing, rRNA, tRNA, and pre-RNA steps. As shown, noncoding QED codons control and regulate transcription and pre-RNA splicing to obtain mRNA. Alternative splicing control allows multiple proteins to be generated from one gene. QED codons translate the mRNA code to synthesize a protein.

(b) the standard protein production path in prokaryotes is from DNA to mRNA to protein production at the ribosome. QED codons translate mRNA like triplet code does at the ribosome.

(c) mRNA is the starting material in viruses, as shown for prokaryotes in (b). Viruses use reverse transcriptase to convert virus mRNA to complementary DNA (cDNA) and use host processing tools to synthesize proteins. QED codons translate mRNA into protein.

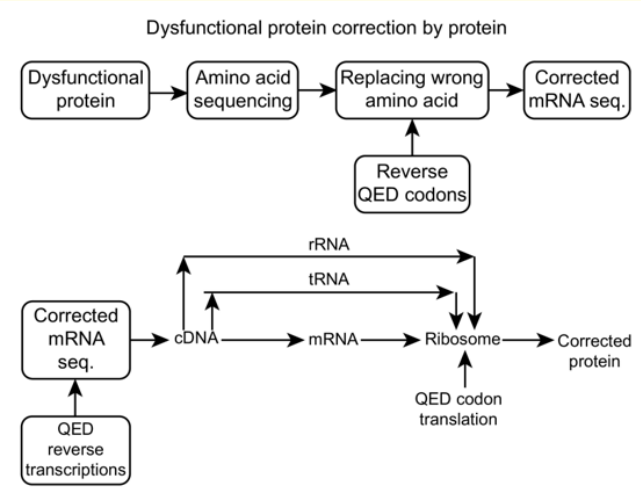

Figure 3: Dysfunctional protein correction at protein level.

Dysfunctional protein is the staring material. Finding the amino acid sequence of the dysfunctional protein is the first step, and the correct amino acid replaces the incorrect amino acid. QED reverse codons generate the correct mRNA. Next, cDNA is generated using QED transcriptase, and QED codons translate into corrected functional proteins.

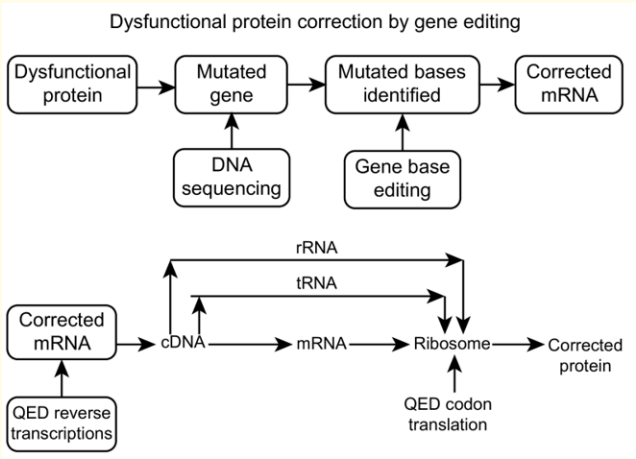

Figure 4: Dysfunctional protein correction at DNA level using gene editing.

Gene Therapy - correction of dysfunctional proteins at the DNA level requires additional steps.

First, the mutated gene is identified, followed by DNA sequencing to identify the mutated bases. A corrected mRNA sequence is generated using CRISPR base-editing tools. The rest of the process for obtaining the corrected protein is outlined in Figure 4.

Summary

The QED genetic coding for eukaryotic cells equally applies to prokaryotes and viruses. The QED codon has a new encoding -protein and a noncoding regulatory codon table. The QED eukaryotic encoding overcomes the triplet codon limitations. The steps for correcting dysfunctional proteins are described. Developing cures for monogenic rare, multigenic cancers and neurodegenerative diseases are envisioned.

Data Availability

N/A.

Code Availability

N/A.

Acknowledgements

I am thankful to Dr. Rakesh Dixit, MD, Tiburon, CA, for discussing diseases and to Dr. Nawin Mishra, Distinguished Professor Emeritus of Genetics and Genomics, University of South Carolina, Columbia, SC 29208, for his encouragement and valuable comments regarding this work. My daughter, Usha Singh's rare, incurable disease was the catalyst to start the lifelong journey to find a gene-disease causality relationship. The outcome is the development of QED genetic coding for eukaryotic cells. This work would never have materialized without the unwavering full support of my family: my wife, Amrita Singh, and my children, Asha Singh and Om Prakash Singh (a.k.a. Tom Singh).

Author Contribution

Rama Shankar Singh - 100%.

Competing Interest

No competing interest.

Additional Information

Correspondence and requests for materials should be addressed to Rama Shankar Singh, email: gen-med @comcast.net.

Bibliography

1. Watson JD and Crick F H. "The structure of DNA". *Cold Spring Harbor Symposia on Quantitative Biology* 18 (1953): 123-131.
2. Watson JD and Crick FHC. "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid". *Nature* 171 (1953): 737-738.
3. Crick FHC., *et al*. "The Nobel Prize in physiology or medicine 1962". NobelPrize.org. Nobel Prize outreach AB (2022).
4. Crick F H., *et al*. "General nature of the genetic code for proteins". *Nature* 192 (1961): 1227-1232.
5. Crick FH. "On the genetic code". *Science* 139 (1963): 461-464.
6. Crick F H. "Codon--anticodon pairing: the wobble hypothesis". *Journal of Molecular Biology* 19 (1966): 548-555.
7. Holley R W., *et al*. "The Nobel Prize in physiology or medicine 1968". NobelPrize.org. Nobel Prize outreach AB (2022).

8. Morgan A R., *et al*. "Studies on polynucleotides, LIX. Further codon assignments from amino Acid incorporations directed by ribopolynucleotides containing repeating trinucleotide sequences". *Proceedings of the National Academy of Sciences of the United States of America* 56 (1966): 1899-1906.

9. Nirenberg M and Leder P. "RNA codewords and protein synthesis. The effect of trinucleotides upon the binding of sRNA to ribosomes". *Science* 145 (1964): 1399-1407.

10. Jones OW and Nirenberg M W. "Qualitative survey of RNA codewords". *Proceedings of the National Academy of Sciences of the United States of America* 48 (1962): 2115-2123.

11. Holley RW., *et al*. "Structure of a ribonucleic acid". *Science* 147 (1965): 1462-1465.

12. Jacob F and Monod J. "Genetic regulatory mechanisms in the synthesis of proteins". *Journal of Molecular Biology* 3 (1961): 318-356.

13. Jacob F., *et al*. "The Nobel Prize in physiology or medicine 1965". NobelPrize.org. Nobel Prize outreach AB (2022).

14. Berget S M., *et al*. "Spliced segments at the 5' terminus of adenovirus 2 late mRNA". *Proceedings of the National Academy of Sciences of the United States of America* 74 (1977): 3171-3175.

15. Manley JL., *et al*. "DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract". *Proceedings of the National Academy of Sciences of the United States of America* 77 (1980): 3855-3859.

16. Roberts R J and Sharp PA. "For their discoveries of split genes The Nobel Prize in physiology or medicine 1993". NobelPrize.org. Nobel Prize outreach AB (2022).

17. Nilsen TW and Graveley B R. "Expansion of the eukaryotic proteome by alternative splicing". *Nature* 463 (2010): 457-463.

18. McManus C J and Graveley BR. "RNA structure and the mechanisms of alternative splicing". *Current Opinion in Genetics and Development* 21 (2011): 373-379.

19. Kornberg R D. "The Nobel Prize in chemistry 2006". NobelPrize.org. Nobel Prize outreach AB (2022).

20. Maston GA., *et al*. "Transcriptional regulatory elements in the human genome". *Annual Review of Genomics and Human Genetics* 7 (2006): 29-59.

21. Novoyatleva T., *et al*. "Pre-mRNA missplicing as a cause of human disease". *Progress in Molecular and Subcellular Biology* 44 (2006): 27-46.

22. Ward A J and Cooper TA. "The pathobiology of splicing". *Journal of Pathology* 220 (2010): 152-163.

23. Palade G E. "A small particulate component of the cytoplasm". *The Journal of Biophysical and Biochemical Cytology* 1 (1955): 59-68.

24. Ramakrishnan V., *et al*. "Nobel Prize in Chemistry "for studies of the structure and function of the ribosome". The Nobel Prize in Chemistry 2009, NobelPrize.org (2009).

25. Wimberly BT., *et al*. "Structure of the 30S ribosomal subunit". *Nature* 407 (2000): 327-339.

26. Selmer M., *et al*. "Structure of the 70S ribosome complexed with mRNA and tRNA". *Science* 313 (2006): 1935–1942.

27. Ogle J M and Ramakrishnan V. "Structural insights into translational fidelity". *Annual Review of Biochemistry* 74 (2005): 129-177.

28. Ramakrishnan V. Gene Machine: The Race to Decipher the Secrets of the Ribosome (Hachette Book Group) (2018).

29. Demeshkina N., *et al*. "A new understanding of the decoding principle on the ribosome". *Nature* 484 (2012): 256-259.

30. Rozov A., *et al*. "Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code". *Nature Communications* 7 (2016): 10457.

31. Carter AP., *et al*. "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics". *Nature* 407 (2000): 340-348.

32. Liu CC and Schultz PG. "Adding new chemistries to the genetic code". *Annual Review of Biochemistry* 79 (2010): 413-444.

33. DeBenedictis E A., *et al*. "Multiplex suppression of four quadruplet codons via tRNA directed evolution". *Nature Communications* 12 (2021): 5706.

34. de la Torre D and Chin JW. "Reprogramming the genetic code". *Nature Reviews Genetics* 22 (2021): 169-184.

35. Kolber NS., *et al*. "Orthogonal translation enables heterologous ribosome engineering in *E. coli*". *Nature Communications* 12 (2021): 599.

36. Malyshev DA., *et al*. "Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet". *Proceedings of the National Academy of Sciences of the United States of America* 109 (2012): 12005-12010.

37. Hoshika S., *et al*. "Hachimoji DNA and RNA: a genetic system with eight building blocks". *Science* 363 (2019): 884-887.

38. ME Sakes., *et al*. "The transfer RNA identity problem: a search for rules". *Science* 263 (1994): 191-197.

39. Agris PF. "Decoding the genome: a modified view". *Nucleic Acids Research* 32 (2004): 223-238.

40. Peabody DS. "Translation Initiation at Non-AUG Triplets in Mammalian Cells". *Journal of Biological Chemistry* 264 (1989): 5031-5035.