Application of a periodic table for the genetic code to influenza A/H3N2 virus

Susumu Morimoto

Institute for Science Education

1-20-3 Suimei, Otsu, Shiga, 520-0531 Japan

If biologists can utilize a table to have access to biological phenomena in the manner analogous to the periodic table for chemical elements, they may get hold of a directing post in life science. Currently the mutational rule of influenza viruses have remained perplexed and to reveal it should be now desired when avian influenza virus has just then threatened human beings. Here I examine the applicability of a novel periodic table for the genetic code to influenza A/H3N2 virus, while presenting two rules regarding single point mutations of its virus hemagglutinin gene. One rule is that non-synonymous single point mutations are intimately associated with the first or second base replacements between four groups (5, 6, 9, and 10) in the periodic table. Another rule is that there is a new index (inversion number) conserving in mutation. This paper will take the first step to such contribution of mutational reactions.

Introduction

I presented a periodic table for the genetic code[1] . On the occasion of examining the utility of the proposed table, I will provide a brief explanation of the influenza A/H3N2 virus hemagglutinin (HA) gene in the beginning.

Influenza viruses belong to the *Orthomyxoviridae* family, and three types, A, B, and C, distinguished by the virus's antigenic proteins, infect humans. Of these three types, the greatest pandemic threat comes from type A strains that infect animals, including pigs,

birds, and humans. The best example is the 1918 “Spanish flu” influenza pandemic which is the worst infectious one in history, resulting in about 50 million deaths worldwide. In contrast, type B viruses cause benign disease and infections from type C viruses are infrequent. This is one of some reasons to focus on the influenza A virus.

The influenza A virus is a negative-stranded RNA virus. The virus is divided into many subtypes based on major differences in the surface proteins hemagglutinin (HA) and neuraminidase (NA). The 16 known subtypes of HA (H1-H16) and nine subtypes of NA (N1-N9) occur in many different combinations, of which, H1N1 and H3N2 have mostly circulated in humans. The HA proteins are responsible for binding to receptors on host cells and initiating infection, whereas the NA proteins, which cleave the terminal sialic acid residue from glycoconjugates, are essential for viral proliferation and infectivity[2-5]. To put it more briefly, the HA proteins are the principal target of the host's immune system. This is a reason to focus on the HA protein.

Influenza viruses outwit their hosts by virtue of the accumulation of base replacements in RNA, including the HA mutational modulation of a benign strain into a malign one, and reduce the efficacy of influenza vaccines available before now. Thus if some rules of the mutation of influenza viruses can be revealed, they might make a significant contribution to effective attenuation of mutated influenza viruses. This is a thread of hope to focus on the influenza A virus.

This paper’s purpose is not to provide a perfect rule of mutational reactions of the influenza A virus HA gene, but devote myself to the applicability of a periodic table for the genetic code to a biological system. Of course, some scientists have presented their respective periodic tables for the genetic code, but most of their tables are involved

in the origin and development of the genetic code[6-8] and I do not know the application of such a table to a real life system. Thus I here focus on single point mutations of the HA genes of the contemporary H3N2 subtype.

Theoretical model

To provide evidence to the approach analogous to a periodic table for the chemical elements, I will initially give a short story of such a novel periodic table for the genetic code (Table 1).

Table 1. A periodic table for the genetic code

Table 1 is made through several steps as follows (original form[1] of this table):

1) Change each of decimal numbers (codon numbers) 0-to-63 into a 6-digit binary number.

2) Assign each of four codonic bases to a 2-digit binary number:

C = 00, U = 11, G = 10, and A = 01. (1)

This assignment is based on a scenario of base-pairing of the standard genetic code[9]: one that there is no base-pair between cytosine (C) and uracil (U), and the other that most are the Watson-Crick base pairs --- guanine–cytosine (GC) and adenine–uracil (AU), but some are the non-Watson-Crick base pairs such as GU and GA. These events can be reproduced by the following conditions: the base pairs can be observed when two 2-digit binary numbers share at least a single common figure, and such base pairs can not be observed when not so, such as bases C and U. Thus formulae (2) is attained and then one of four possible shemes is formulae (1).

{C, U} = {(00), (11)} and {G, A} = {(10), (01)} (2)

3) Transform a 6-digit binary number into one codonic triplet (full details[1] regarding this transformation). For example:

Decimal number	Binary number	codonic triplet
28	011100	CUA

4) Calculate an inversion number (IN) from binary numbers of a codon and its anticodon, which INs are used as an index of mutation in a later section. The calculation to obtain IN of a codon is described as follows: compare each position of the binary codon number with the corresponding position of the binary anticodon number, and count the number of positions that have different numbers. The total of the positions with different numbers is the IN of that pair base. I'll take an example of the CUA triplet:

		Decimal number	Binary number	IN
Codon	CUA	28	011100	5
Anticodon	UAG	39	100111	

5) Make a table, like a periodic table for the elements, to put codons in order of their codon numbers and to ensure periodicity of amino acids encoded by their codons. Consequently some 'congeners' encode the same amino acid in the table and others do not. The former instance includes that four codons of group zero encode the amino acid proline P and accordingly, each position of their codons is occupied by 0P-3, 16P-5, 32P-1, and 48P-3, respectively. where the numbers alongside P represent the codon number (head) and IN (tail). The latter one includes that group four is likewise occupied

by 4H-3, 20Q-5, 36Q-1, and 52H-3.

6) When a solid line is 1 and a broken line is 0, each of the binary codon numbers is translated into a hexagram and hence their hexagrams are diagrammatic representation of binary codon numbers.

7) The finished table (Table 1) is available when the number of the period (I, II, …) and each base of codons are inserted.

Method

The study of mutations of influenza virus genes involves two tasks: first, what variant does a codon mutate into? ; second, which one of candidate variants has priority in mutation? To examine the first task, I here perform a comparative analysis with three HA genes --- A/Guangdong/25/93 [accession number (AC No.) Z46406], A/Guandong/28/94 (AC No. U48442), and A/Wuhan/359/95 (AC No. AF038268) using the HA genes of the H3N2 subtype isolated in 1993 to 1995 (49 strains in total). Using these three strains as a reference in each year is largely because virologists generally suspect that new strains will develop in Asia, especially China[2,10]. The information of these genes is obtained from data bases on the web site of PubMed Nucleotide Query (www.ncbi.nlm.nih.gov/). One of such HA genes, A/Guangdong/25/93, is shown in Table 2.

Table 2. HA gene of A/Guangdong/25/93 (H3N2)

To examine the second task, I use Swanson's codon ring[11], Jungck's genetic anticode[6], and a new index called IN. The first two things will be explained in later paragraphs.

Results and Discussion

What variant does a codon mutate into? (The first task)

A catalogue of mutated codons found in the three years is shown in Appendix . A sample population in each year is 26 (1993), 11 (1994), and 12 (1995) strains. Appendix shows mutations at 122 positions within the limits of research. Most of them are single point mutations (154 cases), some are double points mutations (6 cases), and one is a full mutation (replacement of triplet).

When single point mutations of the first, second, and third position in Appendix are written in Table 1, then Fig. 1-a, -b, and -c are attained, respectively.

Fig. 1. Diagrammatic representation of codon number shift in a single point mutation at the first, second, and third bases (H3N2).

Each line of Fig. 1 shows characteristic features. Specifically, in Fig. 1a, red line and blue line mutations both depict a group of horizontal lines between codons of the same period, and these codons are inside the left half or right half of the period. Their mutation shift factor Δ is 1 or 3, where Δ is the difference between the both codon numbers. In Fig. 1b, red line mutations solely depict a group of oblique lines between codons in the left half and the right half of 16 codons composing two periods, I and II for example, and their mutation shift factors take integer 4 (for right-up lines) or 12 (for right-down lines). In contrast, blue line mutations depict a group of oblique lines or horizontal lines. In these oblique line mutations, the reference and variant codons are in the smallest codon box (2 codons box) straddling two periods such as periods V and VI, as exemplified by reference codon CAG(36Q) and variant codon CUG(44L), whereas in the horizontal line mutations, the two codons concerned are in the left half and the right half of the same period, respectively. Their shift factors are 8 for the oblique line and 4

for the horizontal line. Fig. 1c shows that all single point mutations at the 3rd base exclusively depict a group of vertical lines --- and hence occur between congener codons. This connects to the events of synonymous substitutions of amino acids. Then their mutation shift factors take integer 16 or 48 in red line mutations and 16 or 32 in blue line mutations.

All of these positional features, hence all of mutation shift factors, are independent upon four possible schemes of formulae (2). In other schemes, however, blue line mutations may sometimes exchange their shift factors each other: in Fig.1c, for example, blue line mutations of shift factor 16 in the present scheme may change into that of 32 in different scheme, and at the same time, the mutations of shift factor 32 may change into that of 16. If that is admitted, shift factors of the mutated codon are represented as follows:

$$1\times4^{n-1}, \quad 2\times4^{n-1}, \quad \text{or } 3\times4^{n-1} \qquad (n = 1, 2, 3) \qquad (3)$$

, where n is the position number of a mutated codonic base.

Formulae (3) reproduces all of the observed shift factors in Table 3, including the numbers in parentheses.

Table 3. Summary of mutation shift factors observed in Appendix

Consequently I accept Table 1 as a periodic table for the genetic code which may contribute importantly to life science as the periodic table for the chemical elements has served in the physical/chemical science.

What variant codon has high-priority in mutation? (The second task)

Does a single point mutation favour any combination between codonic bases under a variety of chemical settings? But attaining a solution to this second task is a formidable

challenge, because base replacements depend on a variety of reaction factors such as the polarity and hydropathy (hydrophilicity/hydrophobicity) of amino acids/ nucleotides, the ionic strength of environmental chemicals, and furthermore viral genetics[12]: variant bases themselves may exert a dominant influence on the sequential base changes. Each organ of a living system certainly differs in its chemical surroundings, but there is a commonly observed feature that all reactions occur in aqueous solutions.

Thus for the second task I discuss three clues to seem useful as information on the susceptibilities of base replacements, instead of their chemical surroundings.

As a prologue of the second task, I investigate two subjects required in a later section. Firstly, I examine to what extent the mutation at the third base alters amino acid. For this purpose, population of alterations of amino acids, and also concomitantly INs used later, is shown in Table 4.

Table 4. Population of alternations of amino acids and INs in single point mutations of the HA gene (H3N2).

Table 4 shows that about 60% of whole-base replacements (93 of 154 cases) occur at the third base. Of those, about 9% (8 of 93 cases) change amino acids. Since amino acid replacements in the influenza A virus HA proteins are crucial to the evolution of the influenza virus[13], this suggests that mutations of first- and second-bases (about 40% of all base replacements) will be centrally responsible for new strains. --- the base replacements at the third position (about 5% of all base replacements) may be ignored in this paper to examine a periodic table.

Secondly, I confirm a familiar rule: mutations favour base replacements between adenine(A) and guanine(G), or cytosine(C) and uracil(U) than between other bases,

regardless of the position of the base; that is, its rule argues that red line mutations occur in preference to blue line mutations. Indeed, Table 5 shows that the population of red line mutations is approximately three times that of blue line mutations.

Table 5. Population of codonic base-combinations in mutations (H3N2).

This result corresponds nearly to the conventional view of mutation --- 'transition' has precedence over 'transversion'[14]. This familiar view is a case of general rule that two which are alike bear a chemically similar trait to each other, such as function, activity or structure: on this occasion, a base will be substituted for a 'like' base when base replacements occur. However, a chemically similar trait is not necessarily identical to a biologically similar trait: life systems will not accept random replacements of the codonic base to sustain their life: that is to say, influenza virus will set some limits on base replacements/evolution, because random replacements of the codonic bases in the virus gene will lead to anharmonicity/malfunction of the amino-acid sequence of the virus proteins. The translation errors of genetic codes in the influenza RNA virus have to be minimized under a given condition.

Thus, to examine the relationship between this translation errors minimization and base replacements in the influenza HA gene is the first clue of the second task. Swanson proposed a codon ring to show the minimized translation errors of the genetic codes[11] and I linked it to the group number in the proposed table[1]. When single point mutations of the first and second bases in Appendix are written on Swanson's codon ring, then Fig. 2 is attained. Disregarding description of mutation of the third position base is based only on the first result of the second task prologue: most of mutations at the third position base produce the synonymous substitution (see Table 4),

hence will exert minor effects on the amino acid sequence of the virus proteins.

Fig. 2. The first and second base replacements on Swanson's codon ring (H3N2).

In this error-resistant binary encoding scheme on the basis of the principle of minimum change coding (a Gray code), amino acids which are alike have codons which are alike and the circular path from a given codon in Fig. 2 describes a route of minimum differences through the 64 codons. The path is a closed loop.

Indeed, four codons sandwitched between two broken lines in Fig. 2 are members of a congener group in Table 1 (the minimum number of their codon numbers is the group number); ' like codons' in the same congener group are located in a nearby region on the codon ring. You can remember here that many of the congeners code for the same amino acid from the build-up principle of Table 1. For example, four codon numbers 44, 28, 60, and 12 located just before the 6 o'clock position belong to group 12 and all of them code for leucine (see Table 1). From these considerations, although some groups contain two kinds of amino acids, such as groups 5 and 6, I put forward a hypothesis as follows:

Hypothesis Single point mutation under the minimum translation errors will occur preferentially between the nearest neighbour codons on the codon ring in order to alleviate potential anharmonicity of the amino-acid sequence of the virus proteins, and as distance between two codons on the codon ring become longer, the mutation between them will have lower probability, because longer distance impairs the minimum translation errors.

Here I firstly pick up group 5 to apply this hypothesis and then examine the dependence of mutational probability upon distance between two codons concerned.

If you examine mutation between codons belonging to group 5, that gives you the third

base replacements, such as red line mutations (synonymous mutation) between codon numbers 5 and 53, or 21 and 37. Of course, such replacements occasionally give you blue line mutations (including rarely non-synonymous mutation) between codon numbers 5 and 21, or 5 and 37, for example. These mutations can be, in fact, observed in Appendix, while including a unfound blue line mutation between codon numbers 37 and 53.

Red line mutations at the first and second bases of group 5 are given by combination with clockwise and counter-clockwise neighbouring groups 6 and 9, respectively. The former examples include mutations between codon numbers 5 and 6, or 53 and 54, and the latter ones include mutations between codon numbers 37 and 41, or 21 and 45. These red line mutations draw some sets of concentric arches in Fig. 2. Other combination between codons in the two neighbouring groups leads to double/triple points mutations. This holds true for combination between codons in the other two groups.

As just described, the red line mutations at the first position base, which usually produce non-synonymous mutation, are provided by two families comprising a total of eight sets of combination of two neighbouring groups: one family has groups 6 and 5, 9 and 10, 2 and 1, and 13 and 14; the other has groups 4 and 7, 11 and 8, 0 and 3, and 15 and 12. The former is on the left half of the codon ring and their shift factor is 1, whereas the latter is on the right half of the ring and their shift factor is 3.

At the same time, the red line mutations at the second position base, which also produce non-synonymous mutation, are provided by two families. The two families are on the upper and lower halves of the codon ring, and their shift factors are 4 and 12, respectively. Each family is moreover divided into two sub-families of the right and

left halves of the ring. These sub-families draw a total of four sets of concentric arches in Fig.2. The four sub-families' members are: the first is groups 5, 9, 6, and 10 (termed the upper-left in connection with a region of the codon ring); the second, groups 7, 11, 4, and 8 (termed the upper-right); the third, groups 1, 13, 2, and 14 (termed the lower-left); the fourth, groups 3, 15, 0, and 12 (termed the lower-right). Of these group combinations, some can be observed, but others not.

Therefore, a controversial matter is the mutational population of each location in the codon ring. Table 6 shows the dependence of mutational population upon on-ring-distance between two codons concerned.

Table 6. Dependence of mutational population upon on-codon ring-distance (H3N2).

Table 6 shows that the third base replacements are generally on track for expectation in the hypothesis, but the first and second bases replacements are not necessarily on the track. Specifically, in the third base replacements, all of the red line mutations have the minimum distance 1 (mutations between the neighbouring codons), but many of blue line mutations prefer the second-best distance 2. In contrast, in the first and second base replacements, red line mutations prefer to have the second- and third-best distance (3 and 5, respectively) than the minimum distance. In particular, that is notable in the second base replacements. This disagreement beween the mutation expected from the hypothesis and the observed one will be conceivably due to a slight mismatch between mathematical/formal description and actual function of nucleotides or amino acids.

Consequently, from the first clue, it turns out that some codons locating at the upper-left of the codon ring have larger mutational populations and , in the context of

the periodic table, these codons lie in the groups 5, 6, 9, and 10 in the periodic table, Table 1.

For the second clue of the second task, description to get closer to the context of chemical settings will be expected. So, I examine the polarity of amino acids instead of them, because when an amino acid of the HA protein is replaced with other one, it reacts under aqueous environments and then its polarity should play an important role.

Jungck reported that the genetic anticode could be organized by the nucleotide hydrophilicities[6]; he made a 4-by-4 boxes table of the genetic anticode through the systematic arrangement of mononucleotides and dinucleotides (of the second and third positions) of anticodons. I showed that each box of the Jungck's table was composed of congener members of one group in Table 1. When the red line mutations of the first- and second-position bases is exhibited in the Jungck's table, Fig. 3 is attained.

Fig. 3. Red line mutations at the first and second bases in the Jungck's genetic anticode table (H3N2).

In the context of the codon ring, four boxes (groups 0, 3, 12, and 15) of the upper-left in Fig.3 are arranged at four groups of the lower-right in the ring; four boxes (groups 5, 6, 9, and 10) of the lower-right in Fig.3 are arranged at four groups of the upper-left in the ring; Four boxes of the upper-right or lower-left in Fig.3 are arranged at four groups of the same region in the ring (the upper-right (groups 4, 7, 8, and 11) or lower-left (groups 1, 2, 13, and 14), respectively).

From results described above, one can observe that all of the first, second, and third bases replacements have the largest mutational population in the upper-left region of

the codon ring (46%: 51 cases of red line mutations 110 in total, see Table 6) and that its region's seven amino acids are compositely most polar, least bulky, and least hydrophobic amino acids (see Figs. 2 and 3). This means that, in the terms of the periodic table, amino acids encoded by groups 5, 6, 9, and 10 have the largest susceptibility to mutation, and suggests that mutation of the HA gene, thereby creating new strains of influenza virus, will be at least highly affected by strongly polar conditions including a water environment and ionic strength. Because most polar, least bulky, and least hydrophobic amino acids --- those having largest value of Woese et al.'s polar requirement --- have precedence over less polar requirement in these mutational reactions; the stronger the molecular polarity, the stronger the molecular coupling with water-environment molecules/ions. This especially will be pivotal in the first and second bases replacements which produce generally non-synonymous mutation --- 92% of their base replacements (56 of 61 cases) alter amino acids (see Table 4).

All of most polar, least bulky, and least hydrophobic amino acids, however, do not gain ascendancy over least polar, most bulky, and most hydrophobic amino acids in mutation, because groups 4 and 8 encoding amino acids H, Q, and R, whose groups belong to the upper-right in the codon ring, do not have large mutational population (see Table 6 and Figs. 2 and 3), and, furthermore, least polar, most bulky, and most hydrophobic amino acids in the lower-left and lower-right of Table 6a (these are encoded by groups 1, 2, 12, 13, 14, and 15) have relatively large mutational population at the first position base (12%: 13 cases of redline mutations 110 in total). These suggest some factors other than the polarity, size, and hydropathy (hydrophobicity/hydrophilicity) of amino acids are required in the mutation of influenza HA gene. And these findings suggest that most

polar, least bulky, and least hydrophobic amino acids may be necessary, but not sufficient, for mutational reactions of the influenza A virus HA gene and that accurate explanation of these mutational reactions requires the hitherto unknown information of the virus proteins, such as functional harmonicity/anharmonicity. Thus the first and second clues of the second task lead to the following rule:

Rule 1 Base replacements of the influenza A virus HA gene are intimately associated with four groups (5, 6, 9, and 10) in a periodic table for the genetic code, whose triplets encode compositely most polar, least bulky, and least hydrophobic amino acids.

As the third clue of the second task, I report a new index called inversion number (IN) which is likely to provide links to mutational reactions of influenza virus, although it does not incorporate chemical settings. A reason for picking IN is that it was one of the key concepts explaining the structure underlying the genetic code: a periodic table for the genetic code (Table 1) had fundamental periodic patterns of INs, such as (3, 5, 1, 3) (the first and second periods) and (5, 3, 3, 1) (the third and fourth periods), which were generally independent four possible scheme. From this result, I reach an opinion by the following speculation: If IN is a meaningful value in the proposed table, it may be conserved in mutational reactions. Table 4 shows such relationship between INs and single point mutations of the influenza virus HA gene.

Table 4 reveals that when mutations are non-synonymous nucleotide substitution (types ① and ②), INs seem to generally serve as a primary index of mutations, because 77% of such substitution (type ②: 49 of 64 cases) do not change IN although

the remaining 23% (type ①: 15 of 64 cases) change it. In contrast, when mutations are synonymous nucleotide substitution (types ③ and ④), INs do not serve as a primary index of mutations: Of their mutations, 44% (type ④: 40 of 90 cases) do not change IN and the remaining 56% (type ③: 50 of 90 cases) change it. But if you admit that synonymous substitution of amino acids will be higher on the list of priorities than conservation of IN in mutations, particularly controversial are only the mutations of types ① and ②. Of these two types, since all mutations of type ② conserve IN, a cause of a defective index (value 77%) is attributed to mutations of type ①.

As a matter of course, I do not think that such a simple idea will give a complete solution to complex mutations in the influenza virus and can not currently envisage any more suitable indexe/tags for mutational reactions within the framework of the periodic table. However, when recalling that this paper's purpose is whether to what degree a periodic table for the genetic code can apply to mutational reactions, I believe that the proposed table will be on a satisfactory level; I am of the opinion that it would be better to get hold of an index of mutations than to be void of it. If you admit that, conserving IN will be very crucial to the base replacements which change the amino acid of HA proteins, since these INs are independent on four possible schemes. Thus, for the moment, I will add a following rule as a result of the third clue of the second task:

Rule 2 When a mutation occurs at the first or second base position (alters an amino acid), it conserves IN in the mutation.

As just described, a periodic table for the genetic code is not good at everything and accurate, valuable explanation/prediction of the base replacements in the influenza A virus HA gene seems to require different approach including the formation/disruption of chemical bond. However, I infer that if acquaintance of mutations is accumulated and is well incorporated into the periodic table for the genetic code, the proposed table will make important/satisfactory contribution to mutational reactions in molecular biology. This paper will take the first step to such contribution.

Appendix . A catalogue of mutated codons of influenza A virus HA genes isolated in 1993, 1994, and 1995 (H3N2)

Position No*1	8	16	23	25	26	29	30	32	33	40	46	48	56	58
Reference codon*2	AGC	GGU	GAC	AGC	ACG	CUG	UGC	GGA	CAC	ACG	ACG	GAC	ACU	CUG
Codon No./ Amino acid*3	9S	58G	6D	9S	33T	44L	11C	26G	4H	33T	33T	6D	49T	44L
Variant codon*2	AUC	GCU	AAC	AGU	ACA	CUA	UGU	GGG	CAU	ACU	ACA	GAU	ACG	UUG
Codon No./ Amino acid*3	13I	50A	5N	57S	17T	28L	59C	42G	52H	49T	17T	54D	33T	47L
Δ*4	4	8	1	48	16	16	48	16	48	16	16	48	16	3
Mutated position of codon*5	2nd	2nd	1st	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	1st
Variant codon*2			GAU											
Codon No./ Amino acid*3			54D											
Δ*4			48											
Mutated position of codon*5			3rd											
*6 1993			○	○			○	○	○	○	○		○	○
1994	○	○				○								
1995					○							○		

59	63	65	66	69	69	73	73	75	77	78	79	86	91	94	96	97	99
GUU	UCA	GGU	AGA	GAC	GAC	CGA	CGA	CUU	GGA	AAA	AAC	CUA	CAU	GGC	CAA	AAU	GAA
62V	19S	58G	25R	6D	6D	24R	24R	60L	26G	21K	5N	28L	52H	10G	20Q	53N	22E
AUU	CCA	AGU	AAA	GGC	AAC	AGA	CGG	CUC	GGG	AAG	AAU	CUU	AAU	GGA	CAG	AAC	AAA
61I	16P	57S	21K	10G	5N	25R	40R	12L	42G	37K	53N	60L	53N	26G	36Q	5N	21K

1	3	1	4	4	1	1	16	48	16	16	48	32	1	16	16	48	1
1st	1st	1st	2nd	2nd	1st	1st	3rd	3rd	3rd	3rd	3rd	3rd	1st	3rd	3rd	3rd	1st

		○	○	○		○		○	○		○	○	○	○	○	○	○
	○				○		○			○							
○																	
105	108	112	115	119	120	121	122	125	126	127	131	137	138	139	140	144	144
GAA	GAA	AGC	CCU	CCG	GAU	UAU	GCU	AGG	UCC	CUA	UCC	AUC	AAU	GAA	GAC	ACU	ACU
22E	22E	9S	48P	32P	54D	55Y	50A	41R	3S	28L	3S	13I	53N	22E	6D	49T	49T

GAG	AAA	AAC	CCC	CCA	GAC	UAC	GCC	AGA	UCA	CUG	UCA	ACC	AAC	GAG	GGC	ACC	GCU
38E	21K	5N	0P	16P	6D	7Y	2A	25R	19S	44L	19S	1T	5N	38E	10G	1T	50A
16	1	4	48	16	48	48	48	16	16	16	16	12	48	16	4	48	1
3rd	1st	2nd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	2nd	3rd	3rd	2nd	3rd	1st

		AAU				CAU						GUC			AAC		
		53N				52H						14V			5N		
		44				3						1			1		
		II				1st						1st			1st		

○	○	○	○	○	○	○		○	○			○					
		○			○							○	○		○	○	
					○		○			○	○			○			○
145	147	149	150	151	151	154	154	157	161	161	161	162	170	172	173	179	182
GGA	GCU	GAU	GGA	AAA	ACA	GCU	GCU	AGG	AAC	AAC	AAA	AGU	UUG	AAA	UUA	GCG	GUG
26G	50A	54D	26G	21K	17T	50A	50A	41R	5N	5N	21K	57S	47L	21K	31L	34A	46V

GGG	ACU	AAU	GGG	GGA	AAA	UCU	ACU	AGA	AAA	AAG	AAC	AGC	CUG	GAA	UCA	GCA	GUU
42G	49T	53N	42G	26G	21K	51S	49T	25R	21K	37K	5N	9S	44L	22E	19S	18A	62V
16	1	1	16	5	4	1	1	16	16	32	16	48	3	1	12	16	16
3rd	1st	1st	3rd	II	2nd	1st	1st	3rd	3rd	3rd	3rd	3rd	1st	1st	2nd	3rd	3rd

				GAC						AAA	AAG				UUG		
				6D						21K	37K				47L		
				15						16	16				16		
				II						3rd	3rd				3rd		

		○		○		○		○	○			○	○	○	○		○

							○			○			○			○	
○	○		○		○						○						
188	192	198	199	199	201	202	205	206	207	209	210	210	211	212	212	213	213
GGC	AAA	GUU	CAU	CAC	CCG	AGU	AGU	GAC	CAG	AGC	CUA	AUA	UAU	GUU	GUU	CGA	CGA
10G	21K	62V	52H	4H	32P	57S	57S	6D	36Q	9S	28L	29I	55Y	62V	62V	24R	24R
GAC	AAG	GUG	CAC	CAU	CCC	AGC	AGA	GAA	CAA	AGU	CUG	CUA	UAC	AUU	GUC	CGC	CAA
6D	37K	46V	4H	52H	0P	9S	25R	22E	20Q	57S	44L	28L	7Y	61I	14V	8R	20Q
4	16	16	48	48	32	48	32	16	16	48	16	1	48	1	48	16	4
2nd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	1st	3rd	1st	3rd	3rd	2nd

○	○	○	○		○		○	○		○	○			○		○	
○			○						○						○		○
		○		○		○						○	○				
217	219	224	230	232	235	235	235	237	239	241	241	242	242	244	247	253	256
AGA	AUA	AGA	AUC	AAU	UCU	UCU	UCU	CCC	GUA	GGU	GGG	CAG	CAG	AGC	GGC	GUA	GGA
25R	29I	25R	13I	53N	51S	51S	51S	0P	30V	58G	42G	36Q	36Q	9S	10G	30V	26G
AAA	ACA	AAA	ACC	GAU	UUU	UUU	UCC	CCA	GUG	GGC	GGU	CUG	AUC	AGU	AGC	GUG	GGU
21K	17T	21K	1T	54D	63F	63F	3S	16P	46V	10G	58G	44L	13I	57S	9S	46V	58G
4	12	4	12	1	12	12	48	16	16	48	16	8	23	48	1	16	32
2nd	2nd	2nd	2nd	1st	2nd	2nd	3rd	3rd	3rd	3rd	3rd	2nd	III	3rd	1st	3rd	3rd
					UCC	UAU					GGC	CGG	CUG				
					3S	55Y					10G	40R	44L				
					48	4					32	4	8				
					3rd	2nd					3rd	2nd	2nd				

○		○	○		○					○		○					○
○	○			○		○		○					○		○	○	
							○		○		○			○			
260	261	262	262	264	264	265	267	271	274	276	278	278	278	281	282	283	284
UUG	AUU	AAU	AAU	ACA	ACA	GGG	CUA	CGG	UUU	AUA	AAU	ACU	AGU	AGC	UCA	AUA	AUG
47L	61I	53N	53N	17T	17T	42G	28L	40R	63F	29I	53N	49T	57S	9S	19S	29I	45M

CUG	GUU	GAU	AAC	ACC	AUA	GGA	UUA	CGA	UUC	CUA	ACU	AAU	AAU	AGU	UCG	AUU	AUU
44L	62V	54D	5N	1T	29I	26G	31L	24R	15F	28L	49T	53N	53N	57S	35S	61I	61I
3	1	1	48	16	12	16	3	16	48	1	4	4	4	48	16	32	16
1st	1st	1st	3rd	3rd	2nd	3rd	1st	3rd	3rd	1st	2nd	2nd	2nd	3rd	3rd	3rd	3rd

	AUC
	13I
	48
	3rd

		○		○		○					○			○	○		
○	○		○				○	○	○	○		○				○	○
					○			○					○				

285	286	289	291	291	292	294	294	296	300	304	313	315	317	320	321	324	331
AGG	UCA	CCC	GGC	GGC	AAC	AGU	AAU	GAA	CCA	AUU	GUA	AAG	ACA	GCC	UGC	UAU	AAA
41R	19S	0P	10G	10G	5N	57S	53N	22E	16P	61I	30V	37K	17T	2A	11C	55Y	21K

AGA	UCG	UCC	GGA	GAC	ACC	AAU	AAA	GGA	CCC	AUC	UUA	AGG	ACU	GCA	UGU	UAC	AAG
25R	35S	3S	26G	6D	1T	53N	21K	26G	0P	13I	31L	41R	49T	18A	59C	7Y	37K
16	16	3	16	4	4	4	32	4	16	48	1	4	32	16	48	48	16
3rd	3rd	1st	3rd	2nd	2nd	2nd	3rd	2nd	3rd	3rd	1st	2nd	3rd	3rd	3rd	3rd	3rd

					ACA	AAA						AGA					
					17T	21K						25R					
					12	36						12					
					II	II						II					

	○		○		○			○		○	○	○		○	○	○	○
○	○	○		○		○			○				○	○	○		○
○				○			○							○			○

333	351	354
GCA	AUC	UUC
18A	13I	15F

GCC	AUU	UUU
2A	61I	63F
16	48	48
3rd	3rd	3rd

○ ○ ○
○

*1, meaning of position numbers of the base triplets as in Table 2.

*2, reference strain in each year is A/Guangdong/25/93, A/Guandong/28/94, and A/Wuhan/359/95.

*3, meaning of the representation of one letter codes including numbers as in Table 1.

*4, Δ = larger value minus smaller value in variant and reference codon numbers, termed mutation shift factor.

*5, II represents a double point mutation and III represents a triplet replacement.

*6, 1993: Florence/1/93 (U49722), Shangdong/9/93 (L76037, Z46417),

Finland/263,274, 276, 278, 292, 295, 296/93 (L75982 ~ 75988),

England/1, 247, 269, 284, 286, 289, 328, 346, 347, 471 /93 (Z46393 ~ 46402),

Scotland/2, 142, 160, 174, 173/93 (Z46412 ~ 46415), Mardrid/252/93 (Z46411)

1994: Sendai/c182, c384/94 (U48439, U48440), Hebei/41/94 (U48441), Akita/1/94 (U48443), Aichi/69/94 (U48446), Saga/447/94 (U65552), Hong Kong/1, 2/94 (Z46407, Z46408), England/67, 68, 7,/94 (Z46403 ~ 46405)

1995: Osaka/c1/95 (U65553), Tochigi/44/95 (U65554), Gifu/2/95 (U65555), Kagoshima/10/95 (U65556), Shiga/20/95 (U65557), Ibaraki/1/95 (U65558), Niigata/124/95 (U65559), Miyagi/29/95 (U65560), Finland/339/95 (L75989),

Akita/1/95 (U48444), Sendai/c375/95 (U48445), Hebei/19/95 (U48447), where numberes in parentheses are accession numbers.

References

1. Morimoto, S. A periodic table for genetic codes. *J Math Chem* **32**, 159-200 (2002).

2. Nakajima, K., Nakajima, S. & Sawai, H. *Influenza* Ch.2-6 (PHP, Tokyo, 1998).

3. Reid, A.H., Fanning, T.G., Hultin, J.V. & Taubenberger, J.K.
Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc. Natl. Acad. Sci. U.S.A.* **96**(4), 1651-1656 (1999).

4. Kawaoka, Y. *Influenza Crisis* Ch.1-4 (Shueisha, Tokyo, 2005)

5. Freemantle, M. Keeping one step ahead of the flu.
Chem. & Eng. News. **83**, No. 9, 49-56 (2005).

6. Jungck, J.R. The genetic code as a periodic table. *J. Mol. Evol.* **11**, 211-224 (1978).

7. Bashford, J.D. & Jarvis, P.D. The genetic code as a periodic table: algebraic aspects.
Biosystems **57**, 147-161 (2000).

8. Biro, J.C. *et al.* A common periodic table of codons and amino acids.
Biochem. Bioph Res. Co. **306**, 408-415 (2003)

9. Muto, A. *RNA World* (eds. Osawa, S. & Shimura, Y.; Kodansha, Tokyo, 1990)
p.123.

10. Plotkin, JB, *et al.* Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl. Acad. Sci. USA* **99,** 6263-6268 (2002).

11. Swanson, R. A unifying concept for the amino acid code.
Bull. Math. Biol. **46**, 187-203 (1984).

12. Suarez, D.L. *et al.* Comparisons of highly virulent H5N1 influenza A Viruses

isolated from humans and chickens from Hong Kong.

J. of Virology, **72**, No.8, 6678-6688 (1998).

13. Lee, M-S. & Chen, J.S-E.

Predicting antigenic variants of influenza A/H3N2 viruses,

Emerg. Infect. Dis. **10**, 1385-1390 (2004).

14. Kimura, M. *Introduction to Molecular Evolution* (ed. Kimura, M.) Ch.1 (Baihukan, Tokyo, 1992).

Table 1. A novel periodic table for the genetic code

		1st base					1st base				
	2nd	C	A	G	U	2nd	C	A	G	U	3rd
I II	C G	0P-3 8R-3	1T-5 9S-5	2A-1 10G-1	3S-3 11C-3	A U	4H-3 12L-3	5N-5 13I-5	6D-1 14V-1	7Y-3 15F-3	C
III IV	C G	16P-5 24R-4	17T-3 25R-3	18A-3 26G-3	19S-1 27-	A U	20Q-5 28L-5	21K-3 29I-3	22E-3 30V-2	23- 31L-1	A
V VI	C G	32P-1 40R-1	33T-3 41R-3	34A-3 42G-3	35S-5 43V-5	A U	36Q-1 44L-1	37K-3 45M-3	38E-3 46V-3	39- 47L-5	G
VII VIII	C G	48P-3 56R-3	49T-3 57S-3	50A-3 58G-3	51S-3 59C-3	A U	52H-3 60L-3	53N-3 61I-3	54D-3 62V-3	55Y-3 63F-3	U

Two numbers alongside one letter, 0P-3 for the CCC triplet for example, represent a codon number (head) and inversion number (tail) (for definitions of these two numbers refer to text). A hexagram is diagrammatic representation of a binary codon numbers, where a solid line is 1 and a broken line is 0.

Table 2. HA gene of A/Guangdong/25/93 (H3N2)

Position No.*																			
17	caa	aaa	cuu	ccc	gga	aau	gac	aac	agc	aca	gca	acg	cug	ugc	cug	gga	cac	cau	gca	gug

37	cca	aac	gga	acg	cua	gug	aaa	aca	auc	acg	aau	gau	caa	auu	gaa	gug	acu	aau	gcu	acu
57	gag	cug	guu	cag	agu	ucc	uca	aca	ggu	aga	aua	ugc	gac	agu	ccu	cac	cga	auc	cuu	gau
77	gga	aaa	aac	ugc	aca	cug	aua	gau	gcu	cua	uug	gga	gac	ccu	cau	ugu	gau	ggc	uuc	caa
97	aau	aag	gaa	ugg	gac	cuu	uuu	guu	gaa	cgc	agc	gaa	gcu	uac	agc	agc	ugu	uac	ccu	uau
117	gau	gug	ccg	gau	uau	gcc	ucc	cuu	agg	ucc	cua	guu	gcc	uca	uca	ggc	acc	cug	gag	uuu
137	auc	aau	gaa	gac	uuc	aau	ugg	acu	gga	guc	gcu	cag	gau	ggg	aaa	agc	uau	gcu	ugc	aaa
157	agg	gga	ucu	guu	aac	agu	uuc	uuu	agu	aga	uug	aau	ugg	uug	cac	aaa	uua	gaa	uac	aaa
177	uau	cca	gcg	cug	aac	gug	acu	aug	cca	aac	aau	ggc	aaa	uuu	gac	aaa	uug	uac	auu	ugg
197	ggg	guu	cau	cac	ccg	agc	acg	gac	agu	gac	caa	acc	agc	cua	uau	guu	cga	gca	uca	ggg
217	aga	guc	aca	guc	ucu	acc	aaa	aga	agc	caa	caa	acu	gua	auc	ccg	aau	auc	ggg	ucu	aga
237	ccc	ugg	gua	agg	ggu	cag	ucc	agu	aga	aua	agc	auc	uau	ugg	aca	aua	gua	aaa	ccg	gga
257	gac	aua	cuu	uug	auu	aau	agc	aca	ggg	aau	cua	auu	gcu	ccu	cgg	ggu	uac	uuc	aaa	aua
277	cga	aau	ggg	aaa	agc	uca	aua	aug	agg	uca	gau	gca	ccc	auu	ggc	aac	ugc	agu	ucu	gaa
297	ugc	auc	acu	cca	aau	gga	agc	auu	ccc	aau	gac	aaa	ccu	uuu	caa	aau	gua	aac	aag	auc
317	aca	uau	ggg	gcc	ugc	ccc	aga	uau	guu	aag	caa	aac	acu	cug	aaa	uug	gca	aca	ggg	aug
337	cgg	aau	gua	cca	gag	aaa	caa	acu	aga	ggc	aua	uuc	ggc	gca	auc	gca	ggu	uuc	aua	gaa
357	aau	ggu	ugg	gag	gga	aug	gua													

Position No.*, These numbers represent the position of base triplets.

Table 3. Summary of mutation shift factors observed in Appendix

		Variant base											
		A			G			U			C		
		1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
Reference base	A				1	4	16	2	8	32	1	4	16
	G	1	4	16				1	4	16	(2)	8	32
	U	2	(8)	32	(1)	(4)	16				3	12	48
	C	1	4	16	(2)	(8)	32	3	12	48			

This table indicates, for example, that when the third base U of the reference codon mutates into the third base C of the variant codon, the mutation shift factor is 48 and is situated in column C (3rd)-row U. In the reverse case (the third base C of the reference codon mutates into the third base U of the variant), the shift factor is 48 and is situated in column U (3rd)-row C. Numbers in parentheses are not

found within the limits of research.

Table 4. Population of alternations of amino acids and INs in single point mutations of the HA gene (H3N2).

		Type & Counting **				
	Mutations*	①	②	③	④	Total
the 1st base	Red line	6	15	4	0	25
	Blue line	3	2	1	0	6
The 2nd base	Red line	1	20	0	0	21
	Blue line	0	9	0	0	9
the 3rd base	Red line	0	0	32	32	64
	Blue line	5	3	13	8	29
Total		15	49	50	40	154

* : meaning of red line and blue line mutations as in Fig.1.

Type ①: reference- and variant-codons conserve neither IN nor amino acid (e.g, D-1 → N-5). Type ②: they both conserve exclusively IN, not amino acid (e.g, E-3 → K-3). Type ③: they conserve exclusively amino acid, not IN (e.g, L-1 → L-5). Type ④: they conserve both IN and amino acid (e.g, R-3 → R-3).

Counting**, Counting of type-population in Appendix.

Table 5. Population of codonic base-combinations in mutations (H3N2).

	Mutated position of codon			
Combination of bases	1st	2nd	3rd	Sum

Red line mutations *	C-U	7	7	34	48
	A-G	18	14	30	62
	Sum	25	21	64	110
Blue line mutations *	G-U	2	1	6	9
	C-A	4	5	14	23
	A-U	0	2	6	8
	C-G	0	1	3	4
	Sum	6	9	29	44

* : meaning of red line and blue line mutations as in Fig.1.

Table 6. Dependence of mutational population upon on-codon ring-distance (H3N2).

a. The first base replacements

	Distance on the codon ring						Location on the codon ring Sum
	1	3	5	7	8	24	
Mutation	Population						
Red line	3	3	3	-	-	-	Upper-left 11
	1	1	-	-	-	-	
	-	1	-	-	-	-	Upper-right 1
	-	3	-	-	-	-	Lower-left 7
	1	3	-	-	-	-	
	1	-	1	-	-	-	Lower-right 6
	-	-	1	3	-	-	
Blue line	-	-	-	-	4	2	6
Sum	6	11	5	3	4	2	31

b. The second base replacements

	Distance on the codon ring									Location on the codon ring Sum
	1	3	5	7	9	13	15	16	31	
Mutation	Population									
Red line	1	3	2	1	-	-	-	-	-	Upper-left 12
	-	-	-	-	4	1	-	-	-	
	-	-	-	-	-	1	1	-	-	Upper-right 2
	-	2	-	2	-	-	-	-	-	Lower-left

										4
	-	1	2	-	-	-	-	-	-	Lower-right 3
Blue line	-	-	1	-	-	-	-	6	2	9
Sum	1	6	5	3	4	2	1	5	2	30

c. The third base replacements

Groups	5			6			9			10			4			7			8			11			
Distance	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
Red line/ population	8	-	-	5	-	-	9	-	-	6	-	-	5	-	-	3	-	-	2	-	-	2	-	-	Sum
Location on the codon/ Sum	Upper-left 28												Upper-right 12												40
Blue line/ population	1	3	1	-	1	-	1	-	-	1	3	1	-	-	-	-	-	-	-	1	-	-	-	-	13
Groups	1			2			13			14			0			3			12			15			
Distance	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
Red line/ population	3	-	-	2	-	-	3	-	-	3	-	-	2	-	-	4	-	-	4	-	-	3	-	-	Sum
Location on the codon/ Sum	Lower-left 11												Lower-right 13												24
Blue line/ population	2	2	-	-	2	-	1	1	-	-	2	-	-	2	1	-	2	-	1	-	-	-	-	-	16

The on-ring-distance is defined by the number of circular arc segment lying between two codons concerned, and hence the distance between two adjacent codons is 1.

The terms 'upper-left, upper-right, lower-left, and lower-right' shows the location of codons on the codon ring.

Captions of Figures

Fig. 1. Diagrammatic representation of codon number shift in a single point mutation at the first, second, and third bases (H3N2).

Two end-points of a line represent two base triplets of base replacement. Then, red

lines show mutations between purine bases or pyrimidine bases, whereas blue lines show mutations between purine and pyrimidine bases.

a, the first base.

b, the second base.

c, the third base.

Fig. 2. The first- and second base replacements on Swanson's codon ring (H3N2).

Values on codon ring are codon numbers. Meaning of red and blue lines as in Fig. 1. Two end-points of the arch line represent two base triplets of base replacement. Four codons sandwitched between two broken lines are congener members belonging to one of 16 groups (groups 0-to-15) in Table 1.

Words U.L., U.R., L.L., and L.R., the denotation of upper-left, upper-right, lower-left, and lower-right, respectively, represent segments of a circle.

Fig. 3. Red line mutations at the first and second bases in the Jungck's genetic anticode table (H3N2).

Mononucleotide relative hydrophilicity increase from the left toward the right of the table, and concurrently from the top toward the bottom of the same table. The number next to each amino acid is its respective value of Woese et al.'s polar requirement. Then Woese's amino acid polar requirement generally increases from the left toward the right of the table.

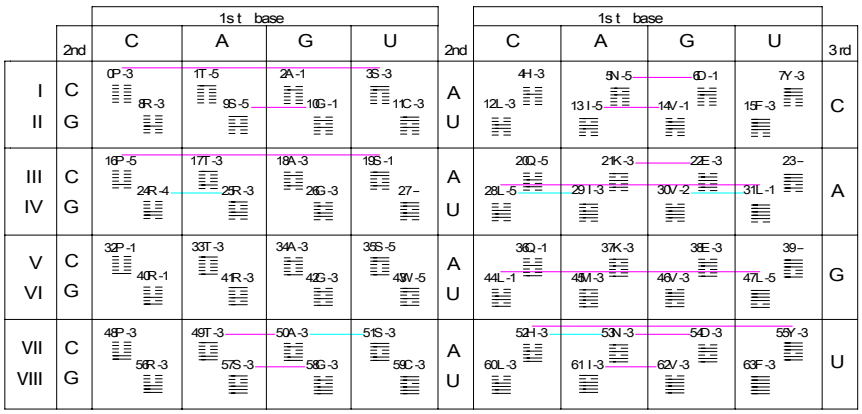
The nine amino acids in six boxes of the groups 4, 5, 6, 8, 9, and 10 are compositely most polar, least bulky, and least hydrophobic amino acids (the minimum number of each box is the group number in Table 1). And the eleven amino acids in nine boxes of the groups 0, 1, 2, 7, 11, 12, 13, 14, and 15 are least polar, most bulky and most

hydrophobic amino acids.

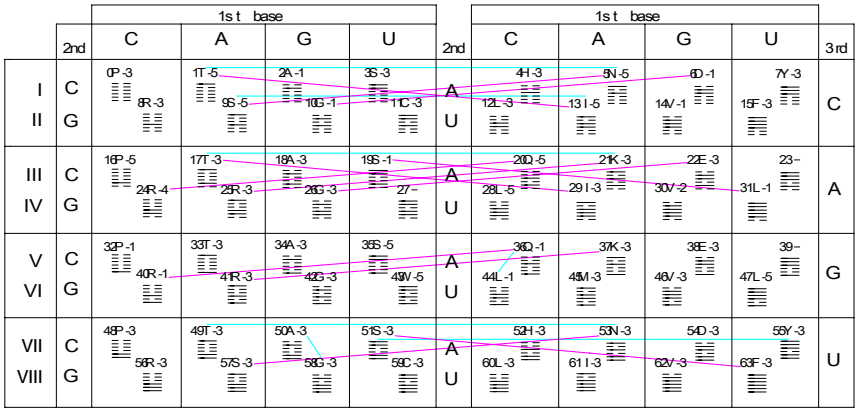
The line 63-51 represents a base replacement between codon numbers 63 and 51, for example. Vertical, red lines show mutations at the first base, whereas horizontal, green lines are those at the second base. Anti-c*, Anticodon

Fig. 1 Diagrammatic representation of codon number shift in a single point mutation at the first, second, and third bases (H3N2).

a. the first base



b. the second base



c. the third base

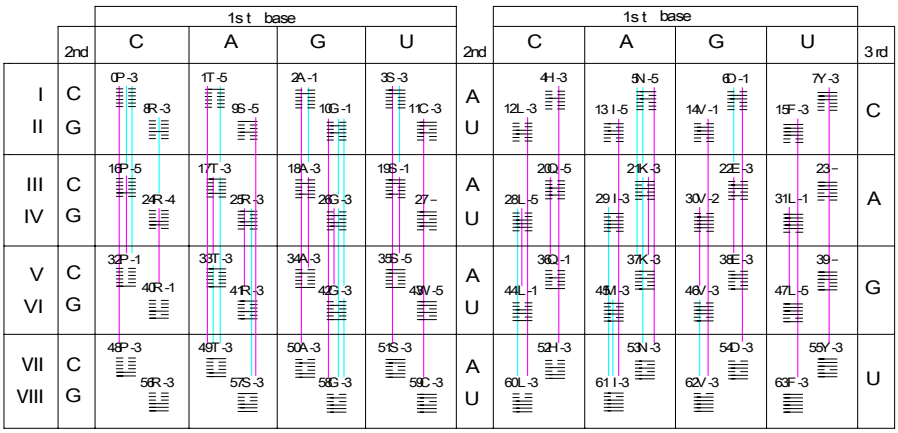


Fig. 2. The first and second base replacements on Swanson's codon ring (H3N2).

a, the first base. **b**, the second base.

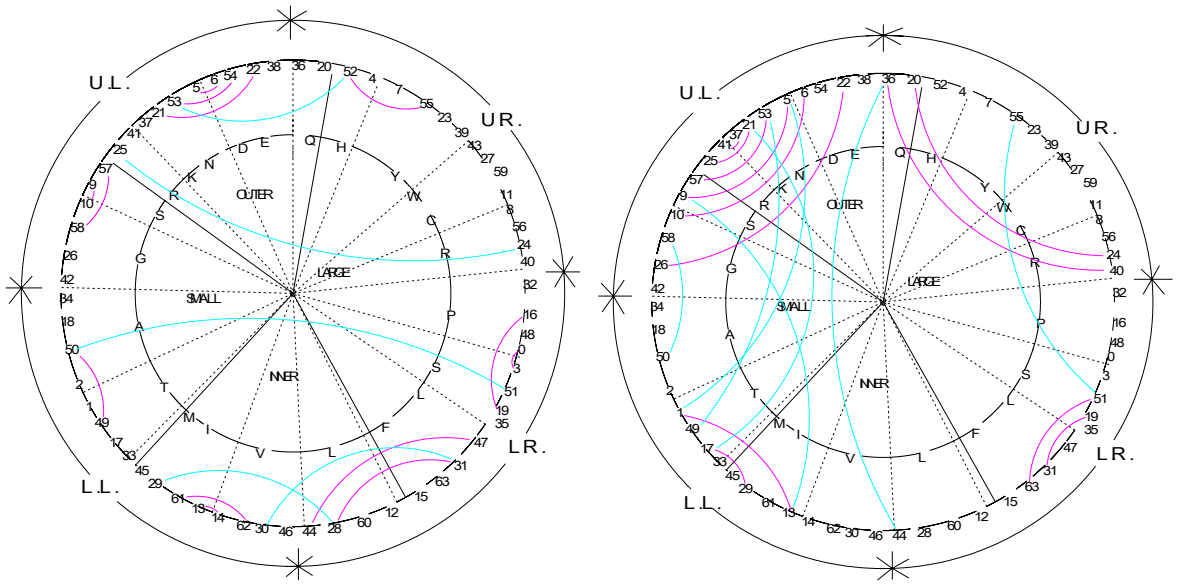


Fig. 3. Red line mutations at the first and second bases in the Jungck's genetic anticode table (H3N2).

Anti-c* 1st base	Anticodon 2nd base: U — Anti-c* /Codon Number	Amino acid	Woese's Polarity	G — Anti-c* /Codon Number	Amino acid	Woese's Polarity	C — Anti-c.* /Codon Number	Amino acid	Woese's Polarity	U — Anti-c.* /Codon Number	Amino acid	Woese's Polarity	Anti-c.* 3rd base
A	21(63)	F	5.0	25 (51)	S	7.5	17(59)	C	4.8	29 (55)	Y	5.4	A
G	22(15)			26 (3)			18(11)			30 (7)			
C	20 (47)	L	4.9	24 (35)			16(43)	W	5.2	28(39)	Stop		

U	23 (31)			27(19)			19(27)	Stop		31(23)	Stop		
A	37 (60)	L	4.9	41(48)	P	6.6	33(56)	R	9.1	45 (52)	H	8.4	G
G	38 (12)			42(0)			34(8)			46 (4)			
C	36 (44)			40(32)			32(40)			44(36)	Q	8.6	
U	39 (28)			43(16)			35(24)			47(20)			
A	5 (62)	V	5.6	9 (50)	A	7.0	1 (58)	G	7.9	13(54)	D	13.0	C
G	6 (14)			10(2)			2 (10)			14 (6)			
C	4 (46)			8 (34)			0 (42)			12(38)	E	12.5	
U	7 (30)			11(18)			3 (26)			15(22)			
A	53 (61)	I	4.9	57(49)	T	6.6	49(57)	S	7.5	61(53)	N	10.0	U
G	54 (13)			58(1)			50(9)			62 (5)			
C	52 (45)	M	5.3	56(33)			48(41)	R	9.1	60(37)	K	10.1	
U	55(29)	I		59(17)			51(25)			63(21)			