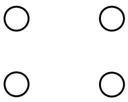
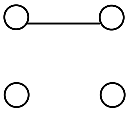
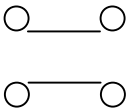
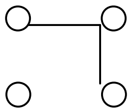
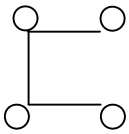
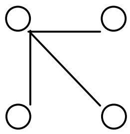
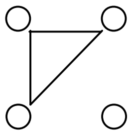
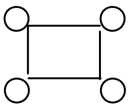
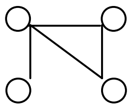
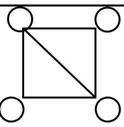
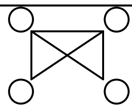
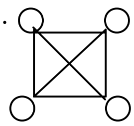
Sd-effect appearing in the genetic code, as the structural explanation of 20 triplets instead of 64

Tamás T.Dénes mathematician
(email: tdenest@freemail.hu)

Nature has built its genetic code on ***4 bases***, these are ***A=adenine***, ***T=thymine***, ***G=guanine***, ***C=cytosine***. The genetic alphabet is built from structures of three of these bases: triplets. As the three spaces in each triplet are made up from one of these 4 bases, so does the number of its combinations based on a simple mathematical formula: $4^3 = 64$.
Curiously, nature has only used just over *20* of this *64* letter alphabet to write the book of living beings[1]. For today's science this is an empirical fact which shows that nature often cannot be described in quantitative mathematical terms.
We will show below that the *Structure-difference effect*[2] (*Sd-effect*) applied for structures defined by four elements (graphs) leads to the exact solution followed by nature itself.

Consider all non isomorphic undirected graphs[3] of 4 vertices (see figures 1.1.-1.12.).

0 edge graph	figure 1.1. 		
1 edge graph	figure 1.2. 		
2 edges graph	figure 1.3. 	figure 1.4. 	
3 edges graph	figure 1.5. 	figure 1.6. 	figure 1.7. 
4 edges graph	figure 1.8. 	figure 1.9. 	
5 edges graph	figure 1.10. 	figure 1.11. 	
6 edges graph	figure 1.12. 		

[1] Interestingly, the most economical language in the world: English uses 26 letters in its alphabet too.
[2] It is the fundamental theorem of the present author's *multistructure theory*.
[3] The graphs G_1 and G_2 are ***isomorphic*** if there is a one to one correspondence between vertices of G_1 and those of G_2 with the property that the number of edges joining any two vertices of G_1 is equal to the number of edges joining the corresponding vertices of G_2.

Sd = Structure-difference, and if G_1 and G_2 are undirected simple graphs, then

(1) $$Sd(G_1, G_2) = |(G_1 \cup G_2| - |G_1 \cap G_2|$$

Figure 1.1.-1.12. introduces all the 12 possible basic structures of the 4 base molecules. *According to the Sd-effect theorem the triplets are constituted from these 12 basic structures in such a way that the Sd-number of the three basic structures in one triplet (the structure-difference) is minimal.*

LEMMA
Let G_1, G_2, G_3 be non isomorphic undirected simple graphs with *n* vertices, then

(2) $$\boldsymbol{\min (Sd(G_1,G_2)+Sd(G_1,G_3)+Sd(G_2,G_3)) = 4}$$

Proof
As the graphs are not isomorphic, the *Sd-number* of any two of them is at least 1. We have proven that out of three *Sd-number*, at least one is 2.

(3) $$Sd(G_1,G_2)=1 \Rightarrow G_2=G_1 \oplus e$$

(4) $$Sd(G_2,G_3)=1 \Rightarrow G_3=G_2 \oplus f \ \ (f \neq e, \text{ because } e \in G_2)$$

(5) $$G_3 \ominus G_1=G_2 \oplus f \ominus G_1= G_1 \oplus e \oplus f \ominus G_1= e \oplus f \Rightarrow Sd(G_1,G_3)=2$$

Q.E.D.

The next theorem can also be deduced from the above Lemma.

THEOREM 1.
If we choose arbitrary three combinations of 4 point basic structures out of Figures 1.1.-1.12., then the minimum sum of the *Sd-number* of each pair of three structures will be *4.*

----- . -----

The Figures 2.1., 2.2., 2.3., …, 2.24. will show all the possible structure triads (out of the above basic structures) where the sum of *Sd* is *4*.
By doing so, we apply the Sd-effect to the linking structures of base molecules. In other words, our model uses the coding machanism of nature (structural code) rather than quantitative combinatorics. The latter one showing possibilities of theoretical model which results in the question: *If there are 64 possibilities, why does nature only use 20?*

The simple fact is that the simple model of combinatorics is not suitable to describe the creation of triplets. Our structural model based on *Sd-effect*, on the other hand, shows exactly *20 (+4)* triplets, as it is in real life.[4]

The combinatoric model should not be applied for the four „names” of amino acids, but to their linking structures. This explains that the bases of life variety is not 64 triplets but the result of the following combinatorial calculation:

[4] The extra four triplets being the dividing symbols in the code.

The number of all 4 element structures (not isomorph structures)= $2^{2^4} \cdot 4! = 1.572.864$

The triplets are created by choosing three of these.

The number of possible choices = $\binom{2^{2^4} \cdot 4!}{3} = 648.517.109.391.294.464$

which is 108.086.184 times the living population of the world today!

If the possible number of structures in one triplet is so great, we can easily understand the enormous variety in DNA chain where there are tens of thousands of these triplets linked together. This is the combinatoric explanation of the real variety in living beings.

----- . -----

Key to Figures 2.1-2.24:
A=**adenine** (red), T=**thymine** (yellow), G=**guanine** (blue), C=**cytosine** (green)

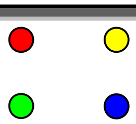
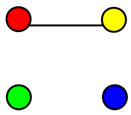
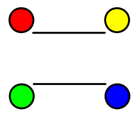
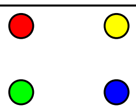
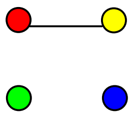
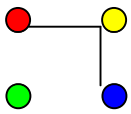
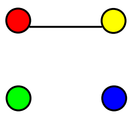
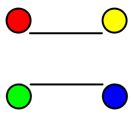
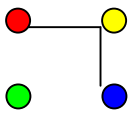
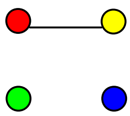
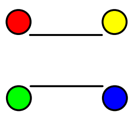
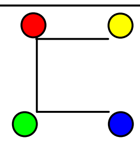
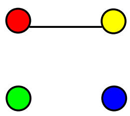
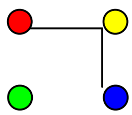
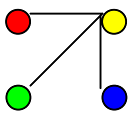
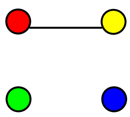
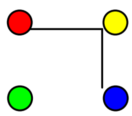
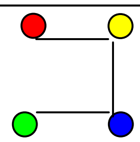
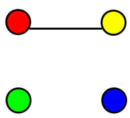
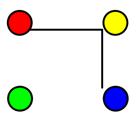
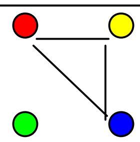
	G_1	G_2	G_3	$\sum$ ***Sd=Sd(G$_1$,G$_2$)+ +Sd(G$_1$,G$_3$)+ +Sd(G$_2$,G$_3$)***
figure 2.1.				$\sum$ **Sd = 4**
figure 2.2.				$\sum$ **Sd = 4**
figure 2.3.				$\sum$ **Sd = 4**
figure 2.4.				$\sum$ **Sd = 4**
figure 2.5.				$\sum$ **Sd = 4**
figue 2.6.				$\sum$ **Sd = 4**
figure 2.7.				$\sum$ **Sd = 4**

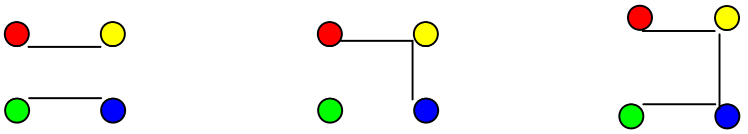
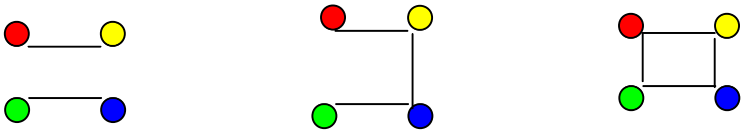
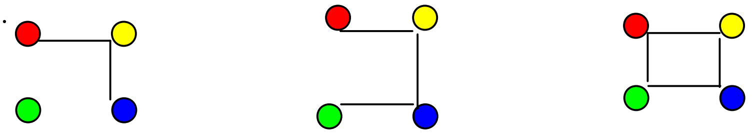
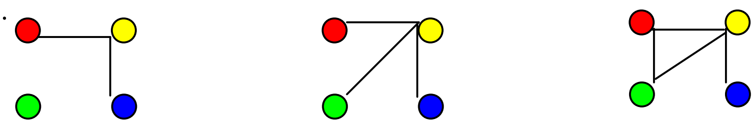
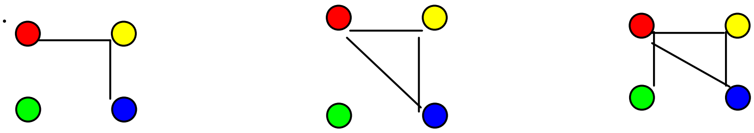
figure 2.8. 	$\sum$ **Sd = 4**
figure 2.9. 	$\sum$ **Sd = 4**
figure 2.10. 	$\sum$ **Sd = 4**
figure 2.11. 	$\sum$ **Sd = 4**
figure 2.12. 	$\sum$ **Sd = 4**

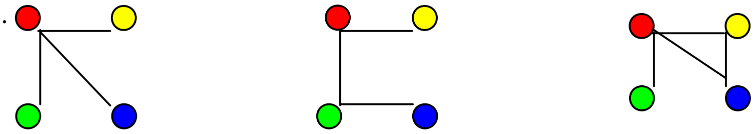
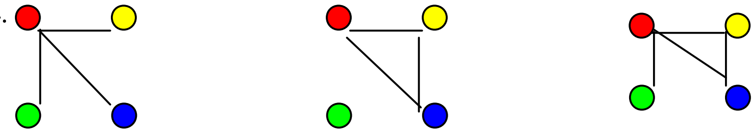
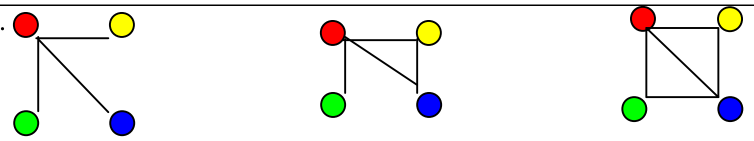
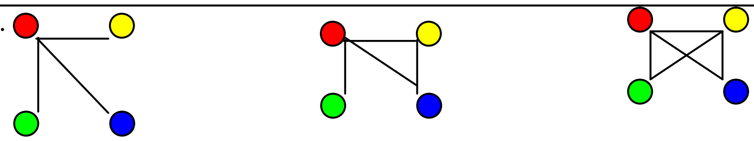
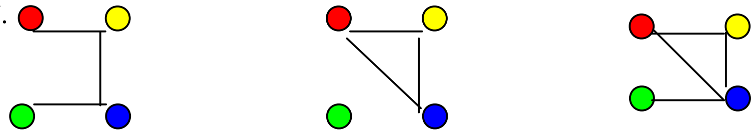
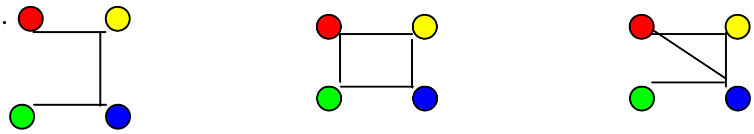
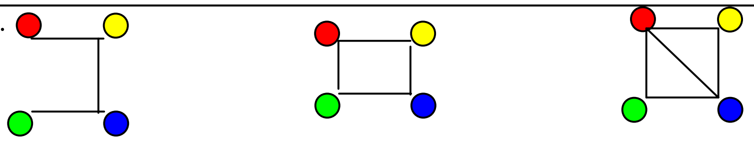
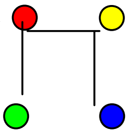
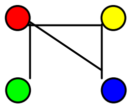
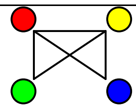
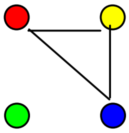
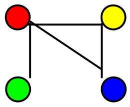
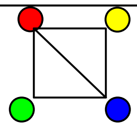
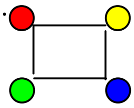
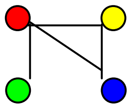
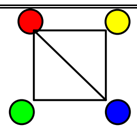
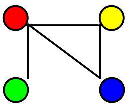
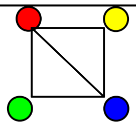
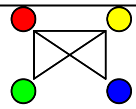
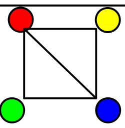
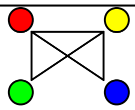
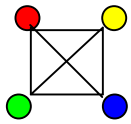
figure 2.13. 	$\sum$ **Sd = 4**
figure 2.14. 	$\sum$ **Sd = 4**
figure 2.15. 	$\sum$ **Sd = 4**
figure 2.16. 	$\sum$ **Sd = 4**
figure 2.17. 	$\sum$ **Sd = 4**
figure 2.18. 	$\sum$ **Sd = 4**
figure 2.19. 	$\sum$ **Sd = 4**

figure 2.20. 				$\sum$ **Sd = 4**
figure 2.21. 				$\sum$ **Sd = 4**
figure 2.22. 				$\sum$ **Sd = 4**
figure 2.23. 				$\sum$ **Sd = 4**
figure 2.24. 				$\sum$ **Sd = 4**

From among the above cases the first and second (see G_1 on fig. 2.1.-2.2.) are degenerate structure, does not appear in the reality because of this. The same one is true for the last G_3 on fig. 2.24. Thus the number of the possible cases are 21, from among which the nature uses one as a sign separating only. Thus it is the structural answer to the title's problem.[5]

Budapest (Hungary), 2007. december 28. Tamás T.Dénes

[5] *„Today it is still an unsolved scientific question why the four bases applied in three spaces in the genetic code define only 20 amino acids instead of 64 in the living material?"*