Recursive Algorithm for the Analysis of Metabolic Cycles

Ovidiu Costica UNGUREANU[1], Marius Constantin POPESCU[2], Elena UNGUREANU [3], Calin Gabriel SARLA[4]

[1, 2, 4] "Vasile Goldis" Western University, Arad, Romania,
[3] "Ion Ionescu de la Brad" University of Agricultural Sciences and Medicine Veterinary, Iasi, Romania.

Summary – This study highlights the implications of the recursive algorithms in the field of biochemistry and genetics for solving the issue of protein synthesis. The existence of some biological structural units congruous with IF statements in a logical schema are emphasised, elaborating a mathematic pattern for the structure of a protein, which is concluded with the drafting of a computer software. The proteins are biomolecules with complex structure, constituting all life forms. The structure of the proteins is insufficiently known due to their dynamicity, the proteins being constantly submitted to processes of synthesis and degradation. In addition, the mathematic pattern can also be used for the analysis of coronavirus which are relatively simple structures, of spherical form covered by peaks formed of nucleoprotein units constituted of nucleic acids, DNA or RNA (as genetic material) and specific proteins. Considering the vital importance for the body proper functioning which these biopolymers have, this study aims to issue simple software which emphasises the various mutations affecting the complex structure of these macromolecular organic compounds, and the mutagenic factors at the basis of the mutations.

Keywords: - *codons, enzymes, proteins, coronavirus, recursion, backtracking*.

1. Introduction

The performed studies underlines that the human genome structure comprises from 23,000 to 30,000 genes, and the speed of genetic mapping is constantly increasing (about 3 genes per day being identified).
Currently, we find higher trend of transferring the information held by genetics from the biochemistry laboratories to the informatics ones where, based on software programs, the genetic sequencing is developed in a personalised manner which leads in the near future to the occurrence of extended personalised medical therapies.
It is slightly probable that there is in the universe another type of molecules with more remarkable properties than the proteins as they provide for the organization and maintaining of the cells morphological structures and also the manifestation of their vital functions and activities which lead to the conclusion that no life forms without proteins existence are known.

Manuscript received September 5, 2020
Manuscript revised September 20, 2020

DOI: 10.22937/IJCSNS.2020.20.09.21

Thus, the chemical reactions in the cells depend on the protein enzymes combination with various substrates. Structures such as those of the muscles depend on the interactions between the proteins, the genetic control is caused by the protein-DNA interactions, the nervous activity implies interactions transmitter – protein, the immune protection relies on antibody-antigen interactions.

The viruses present an extremely simplified structure compared to that of the proteins which cannot ensure them own metabolism, placing them at the border between alive and not alive. This is the reason why a virus in pure state (virion) cannot reproduce alone. For reproduction, the viruses must penetrate inside a specific cell of a living body and trigger the self-replication mechanism through the host cell [1].
If the genomes of some viruses such as varicella and smallpox are formed of DNA (Deoxyribonucleic acid), like human's genome, those of the coronaviruses are formed of RNA. It is known that the RNA (Ribonucleic Acid) viruses have small genomes which change frequently [2]. The coronaviruses, named so due to the distinctive aspect of their peaks (seen at microscope they look like a crown, under which there is a membrane which contains inside the genetic material of the virus/ the genome) are entities with spherical particles of 100-160 nm in diameter, containing a genome with RNA (ssRNA) of 27–32 kb. The genome terminals codify a polyprotein, pp1ab, split in 16 non-structural proteins involved in genome transcription and replication [3].

As there is not complete knowledge on coronaviruses, genes and various processes at genetic level [4], issuing some modern methods of investigation is required, methods relying on rigorous interpretation of the acquired data; this interpretation being carried out exclusively at informatics level. The contribution of computer science allowed for the issuing of various algorithms which may characterise a genetic proves. An important aspect is the applications of the computer science in virology. Thus, various mathematic patterns and computer applications can be built by the analysis of a virus way of action and how it detours the protein synthesis into its favour.

In addition, the mutations help the viruses to adjust and infect new host species. For instance, the belief is that the new coronavirus has its origin in a bats virus but it is still not known what mutations allowed for that leap from animals to humans [5]. This paper is to approach two aspects of a protein structure: physically (the length of the protein unit chain) and chemically (the disposition of the amino acids in a protein and its participation in the good development of a metabolic cycle). The various physical and chemical transformations of a protein structure cause directly the functioning of the respective protein.

Computationally, an analysis of the biological entities (for instance nucleotide, codons, genes), that comprise the information required for protein synthesis is possible.

A protein is formed by the junction of more amino acids by polypeptide bonds, based on some information contained in a DNA molecule. The amino acids represent the non-hydrolysable structural units of the proteins. Though we know in nature about 200 amino acids, only 20-22 of them are used for the proteins biosynthesis and are encoded by the genetic code being also named proteinogenic (ordinary) amino acids. The other amino acids, less spread are called proteinogenic occasional. The lack of amino acids in nutrition causes severe metabolic disorders (deficiency disorders) [6].

The DNA molecule is formed of purine and pyrimidine nitrogenous bases, joined together by phosphodiester bonds. Thus, the DNA molecule represents a chain of nitrogenous bases. A group of three nitrogenous bases forming a unit called a codon. It is known that a certain amino acid corresponds to each codon. The sequence of codons in the structure of the DNA molecule dictates, implicitly, the sequence of amino acids in the structure of proteins. Bodies where metabolic cycles take place are divided into prokaryotes (single-celled organisms without a nucleus) and eukaryotes (bodies that possess a nucleus). The functioning of all living bodies is based on metabolic cycles, which turn the substances in the environment into necessary substances, some absolutely vital to the body. Enzymes are proteins in the structure of a metabolic cycle that intervene in the transformation of various chemicals in its structure. These biocatalysts possess certain special areas called active sites, through which they participate in various biochemical reactions. It is known that certain areas in the structure of a gene undergo mutations with a much higher frequency than other areas, forming a hot spot that can be quickly identified through specialized software programs [7].

If it is possible to determine correctly the structure of a gene, and, implicitly, the structure of a protein (assuming that the protein has a harmful effect on the human body) encoded by this gene, then the gene can be modified by the enzymatic mechanism and by that of reverse transcription.

As for human coronaviruses, there are currently no clinical treatments or prevention strategies for any of them.

One of the successes is the identification of a mutant protein in the blood (which occurs in the case of cervical cancer), long before the actual onset of cancer, by harvesting a tiny amount of blood. The analysis of the protein structure also allows the highlighting of the corresponding gene, as well as the changes in the structure of the gene. Due to structural transformations of functional proteins, large differences occur in metabolic cycles, in which case quite accurate predictions can be made [8]. Due to the large number of genes and the biological processes that take place at their level (the *wobble* concept), for now the matter can only be partially addressed. In addition, a gene can have various variants called *alleles* (these variants of a single gene can also afford mutations). The very high total number of alleles is an impediment in the development of customized programs.

Future coronavirus studies will likely focus on the biological properties of viruses using virus isolation, reverse genetics, and in vitro and in vivo infection testing. So far, for the study of the new coronavirus, a phylogenetic and evolutionary analysis has been applied to characterize the 2019-CVC virus in order to better understand the transmission dynamics [9].

2. Modelling genomes, genes, proteins and coronavirus

Living bodies have the specific property of being able to synthesize their own proteins from amino acids taken from food or resulting from the enzymatic hydrolysis of food proteins. Proteins have the property of being specific frame, because each organ of the same plant or animal contains specific proteins, different from the proteins of other organs of the same individual. Proteins are also species-specific, because the same organ from different species, animal or plant, contains specific proteins, different from the same organ of an individual of another species. The specificity of proteins is also manifested by their immunological properties: the inoculation of a foreign protein in the body of an animal causes the occurrence in its serum of a substance able of precipitating only the protein that has been inoculated. Inoculated substances are called antigens and can be proteins, poly-carbohydrates, complex carbohydrate-lipid-polyprotein associations, which are foreign to the body which they penetrated and consequently trigger the biosynthesis of some specific defence proteins called antibodies. The antigen reacts with the formed antibodies, causing the antigen-antibody reaction, which annihilates the harmful action of the antigen.

The formation of antibodies coincides with the installation in the body of a specific resistance (immunity to various viruses) to the pathogen. Immunological reactions are the basis for the preparation and use of serums and vaccines to immunize bodies against the occurrence and spread of infectious diseases caused by any type of virus [10].

Viruses are classified according to the type of nucleic acid contained in DNA viruses and RNA viruses. Penetrating into cells, viruses divert their own cellular biosynthesis, especially to the synthesis of the molecular components of the virus. Viral RNA and DNA attach to host cell ribosome's with priority over cell nucleic acids and thus biosynthesize viral nucleic acids. What is known for sure is that coronavirus is part of the category of RNA viruses [11]. Due to their composition, consisting of amino acids, proteins are found along with other important compounds such as: polysaccharides, lipids and nucleic acids, starting with the structure of viruses, prokaryotic organisms, eukaryotes and ending with humans [12]. Depending on their role in the body, proteins can be, functional (they are part of the metabolic cycles and are responsible for the body's immune and autoimmune response) and structural (maintain cellular integrity, being found in the cell wall).

The dynamics of a metabolic cycle is coordinated by enzymes, which are included in the category of proteins. Because the total hydrolysis of proteins results in amino acids (structural unit), it is necessary to know the structure of proteins to determine the number and nature of the component amino acids, as well as their binding sequence in the macromolecule. It is then necessary to analyse the configuration of the polypeptide catenary, which has specific internal sizes and structures, which imprint on the protein configurations with different orientations and interactions. At present, the structure of proteins is not completely elucidated, proposing four structural forms. Thus, the primary structure is given by the amino acids that enter the protein chain by forming peptide bonds, -CO-NH- which are in the same plane, and the carbon -CH- can rotate, being able to appear in different planes. The secondary structure refers to the shape and length of the polypeptide chains, properties induced by hydrogen bonds. The most common types of secondary structure are the *alpha, helix and beta chains*. Most of the proteins are helical (spiral), as is α-keratin in hair, myosin in muscle, fibroin in silk. The tertiary structure was highlighted by X-ray crystallography, which proved that protein macromolecules have a three-dimensional conformation, achieved by coupling several short polypeptide chains together, a coupling that leads to the formation of protein fibres. The quaternary structure refers to the way in which protein subunits join. Enzymes that catalyse the assembly of these subunits are called holoenzymes, some of which are called regulatory subunits and catalytic subunits [13]. The arrangement of amino acids in the structure of a protein is conditioned by genetic information in DNA or RNA. Nucleic acids are information biomolecules, having the role of storing and transmitting genetic information. Nucleic acids are important components of all cells, accounting for about 15% of their dry matter content. Chemically, nucleic acids are macromolecular polynucleotides, respectively, DNA acids (DNA) are polydeoxyribonucleotides, and RNA acids (RNA) are polyribonucleotides. The structure of nucleic acids is very similar to the structure of proteins due to their macromolecular nature [14].

Between the different entities that are used in biology and the various mathematical notions, a series of correspondences are identified. Thus, let $A = \{a_1, a_2, ..., a_{20}\}$ be the set of essential amino acids that underlie the formation of a protein. A protein is obtained by chaining several amino acids (for instance, Hexokinase - the first enzyme in the glucose degradation cycle - is made up of 930 amino acids). So we can define a new set $P = \{p_1, p_2, ..., p_n\}$, the set of proteins, which can be made up of amino acids. Thus, there appears a domain of definition (represented by the set A), a codomain (represented by the set P) and a law / method, by which each element from the set A corresponds to an element from the set P ($f(a): A \rightarrow P$). The following properties can be identified [15]:

$$f_1(a_1, a_2, ..., a_{20})=p_1,\ f_2(a_1, a_2, ..., a_{20})= p_2, ...,\ f_n(a_1, a_2, ..., a_{20})=p_n, \quad (1)$$

Where, each protein may have the following form

$$P = a_1+a_2+a_3+...+a_{20} \text{ or } P = \sum_{i=1}^{n} a_1 \sum_{i=1}^{n} a_1 . \quad (2)$$

We may write:

$$\begin{aligned} &f_1(a_1, a_2, ..., a_{20})=a_1+a_2+a_3+...+ a_{20}=p_1,\\ &f_2(a_1, a_2, ..., a_{20})=a_5+a_7+a_{15}+...+a_9=p_2,\\ &.....\\ &f_n(a_1, a_2, ..., a_{20})=a_{10}+a_{11}+a_{14}+...+a_1=p_n. \end{aligned} \quad (3)$$

Summing up the elements that characterize amino acids is not commutative ($a_1 + a_2 \neq a_2 + a_1$), but is associative ($a_1 + (a_2 + a_3) = (a_1 + a_2) + a_3$), because by changing the location of an amino acid in the structure of a protein, its properties will change and very importantly, a new protein will appear. For instance:

$$P_1=a_1+a_2+a_3+a_4+a_5 \text{ and } P_3= a_1+a_3+a_2+a_4+a_5, \quad (4)$$

but $P_1 \neq P_3$.

The two properties are based on biochemical considerations, in the sense that the sequence and nature of amino acids determine the identity of the protein. For instance, for the P_1 protein with the previous structure (P_1), the amino acid a_3 participates in the achievement of protein-specific chemical bonds P_1.

By modifying the site of amino acid a_3, the protein P_2 will be obtained with the structure $P_2 = a_3 + a_1 + a_2 + a_4 + a_5$. By changing a_3 to a_2, the amino acid a_2 will participate in other chemical bonds than those of a_3. Another property is observed which shows that,

although the set of essential amino acids is small (20), the set of proteins that are synthesized on the basis of these amino acids can be considered infinite in terms of practical applications.

The arrangement of amino acids in a protein is done according to a rule dictated by the arrangement of codons in the structure of a DNA or RNA molecule. Thus, each essential amino acid corresponds to a specific codon that is composed of nitrogenous bases: Adenine, Guanine, Cytosine and Thymine (in the case of DNA and uracil, instead of thymine in the RNA molecule).

For instance, the amino acid, isoleucine, corresponds to the CUC codon (*Cytosine-Urcyl-Cytosine*) in which case all the codons in the structure of nucleic acids (DNA, RNA) can be the basis for the formation of a C set.

As an amino acid corresponds to a codon, a function can be defined Ψ:C→A. Therefore, two functions were defined Ψ:C→A şi f:A→P, from where, we can notice that set A is the codomain for the function Ψ and the domain of the function f. For the function Ψ:C→A, the following situation may exist: either codons c_1, c_2, c_3, c_4 which the amino acids a_1, a_2, a_3, a_4 correspond to, or detailing, c_1–AAG–phenylalanine–a_1, c_2–AGU–serine–a_2, c_3–ACG–cysteine–a_3, c_4–UCU–arginine–a_4, and $a_1+a_2+a_3+a_4=P_4$ (a simplified situation, as various biological interpretations such as the concept of "wobble" were not taken into account.").
From the mentioned instance it results that the P4 protein can be composed of codons, namely $P_4=c_1+c_2+c_3+c_4$. One can see a situation similar to that of the composition of functions Ψ: C→A and f: A→P, thus we can define the function Δ: C→P, whereby a protein is defined by the succession of codons (a real biological situation). However, various biological considerations, such as evolutionary pressure, have led to the grouping of codons in a linear structure called gene. Thus, a gene (G_1) with the configuration $G_1=c_1+c_2+c_3+c_4+c_5$, presents a similarity to the structure of a protein ($a_1+a_2+a_3+a_4+a_5=P_1$). The gene $G_1=\Sigma c_i$ contains the information required for the synthesis of a protein and, therefore, a μ function can be defined which has as domain the set of genes in an organism (30,000, in the case of humans), and as codomain, the set of proteins that can be synthesized based on the information contained in these gene (μ:G→P). Although, over time, the definition of a gene, as a "protein", has undergone additions, it has remained valid. The function μ is a bijection $\mu(G_1)=P_1$, $\mu(G_2)=P_2$, … , $\mu(G_n)=P_n$. As such, a first general representation of the process can be made (Fig.1).

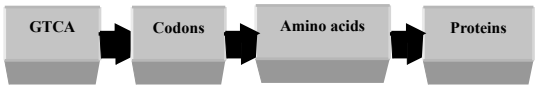

Figure 1. General representation of the process.

It is observed that starting from a set with only a few elements, one arrives at a set with an infinite number of elements (this situation is also the result of biological processes such as, for instance, the mutational process).

2.1 Structural proteins

Structural proteins participate in the proper functioning of some organs. For instance, proteins: actin and myosin play a role in the development of muscle fibre, keratins are constituent proteins of the epidermis, hair, feathers and horn formations: nails, horns, hooves) collagen is in connective tissue, skin, tendons, cartilage, ligaments, elastin is involved in the structure of elastic fibres in arteries and tendons, and fibrinogen is 4% of blood plasma proteins with an essential role in blood clotting.

At the same time, the arrangement of amino acids in the structure of proteins provides information regarding the characterization of the protein according to the number of amino acids, and implicitly the length of the protein chain. For instance, hexokinase has in its structure 930 amino acids (its length, which can be considered to be equal to 930 - L(hk)=930). Usually, proteins have a chain length given by the number of amino acids. It is estimated that there are between 300 and 500 amino acids in the structure of a protein (so it can be written $L(P) \in \in [300, 500]$). Due to the mutational process, the length of a protein can vary.
For instance, if $P_1=a_1+a_2+a_3+a_4+a_5$, so $L(P_1)=5$, then due to the mutations at the level of the gene G_1 (gene conditioning protein P_1), protein P_1 turns into protein P_2, and gene G_1 into gene G_2, also occurring the situation as $L(P_1)\neq L(P_2)$. This process is the basis for the appearance of new species or on the contrary for the extinction of that species. The evolutionary process was observed in nature and can be summarized as a logical scheme ($G_1\rightarrow G_2$, deci $L(P_1)\neq L(P_2)$), which illustrates an analogy between a computer "logical IF" and a biological "IF." (Fig. 2).

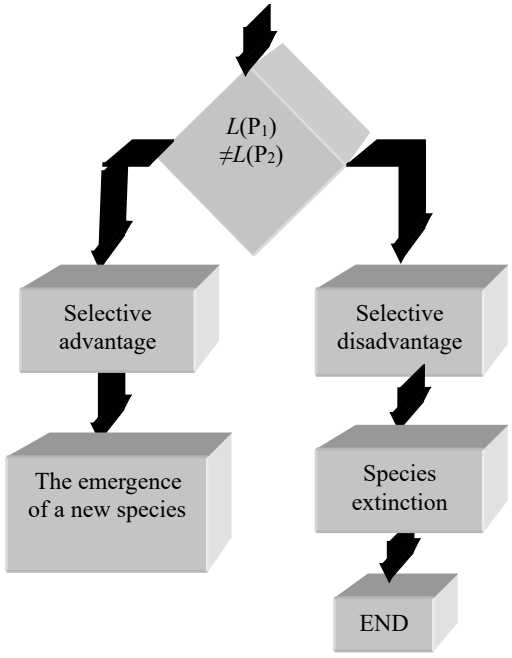

Figure 2. Logical scheme on process evolution.

Taking into account the length of the protein chain (physical property), it refers directly to the structural character of the protein and not to the functional one. Thus, a change in length can have the effect of inactivating various active sites, and the protein can no longer participate in the architecture of the body.

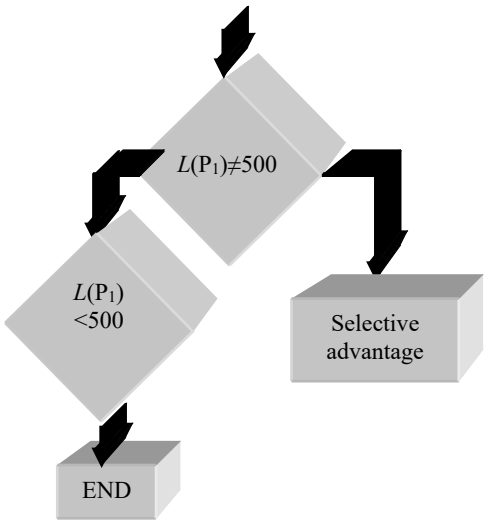

Figure 3. Logical scheme on the process evolution.

For instance, be it a protein with length $L(P)=500$, constituted of 500 amino acids; if its length goes below 500 ($L(P)<500$), then all bodies possessing this length are progressively eliminated compared to the bodies having $L(P)=500$ and which are viable (Fig.3). In biology it is common to find, for instance, 4 proteins with characteristics $P_1=100$ amino acids (i.e., $L(P_1)=100$), $P_2=150$ amino acids (i.e., $L(P_2)=150$), $P_3=180$ amino acids (i.e., $L(P_3)=180$) and $P_4=200$ amino acids (i.e., $L(P_4)=200$), that enter in the formation of a body in the following order $P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow P_4$, with representation from the Fig.4

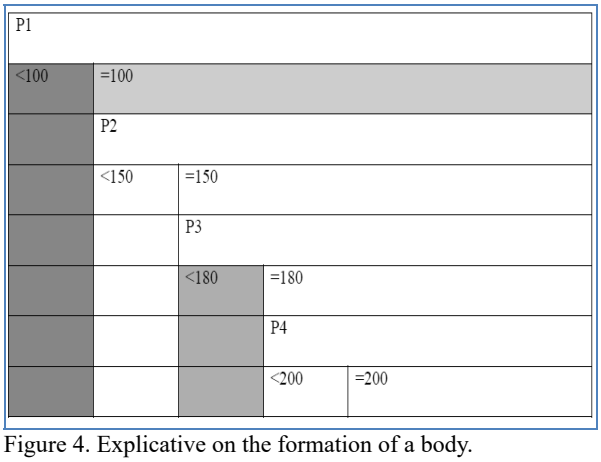

Figure 4. Explicative on the formation of a body.

Genes G_1, G_2, G_3, G_4 correspond to the proteins P_1, P_2, P_3, P_4, which is $G_1 \rightarrow P_1$, $G_2 \rightarrow P_2$, $G_3 \rightarrow P_3$, $G_4 \rightarrow P_4$. There is a possibility of mutations in the genes, which can have the effect of inactivating proteins by changing their length; the result being the appearance of mutant organisms, which are eliminated. By biochemical analyses, the existing proteins can be identified, and the situation can be represented graphically.

2.2 Functional proteins

Functional proteins have an important metabolic role because they are found inside cells and in body fluids. They enter in the constitution of some enzymes (as protein part - oxidoreductases, transferases, ligases and lyases), of some pigments (chloroglobins, haemoglobin). For instance, a set of proteins from the structure of a linear metabolic cycle is considered and the various situations that occur as a result of structural changes are analysed, highlighting various logic schemes that appear from the description of that metabolic cycle. Let the next metabolic cycle:

$$E_1\ E_2\ E_3\ E_4 \quad \text{and} \quad A \rightarrow B \rightarrow C \rightarrow D \rightarrow X,$$

where: E_1, E_2, E_3, E_4 – are enzymes (originally proteins), which act at various stages of the metabolic cycle; A - the initial substance subject to transformation; B, C, D - the intermediate substances of the metabolic cycle (represent the substances on which the enzymes of the cycle act); X - the final synthesized substance (the substance resulting from the completion of the entire metabolic cycle).

The respective enzymes have the following properties:

$$L(E_1)=50,\ L(E_2)=100,\ L(E_3)=110,\ L(E_4)=150,$$

where, E_1 has as active site the domain S_1 formed of the amino acids a_{15}, a_{16}, a_{17}, thus $S_1=a_{15}+a_{16}+a_{17}$; E_2 has as active site the domain S_2 formed of the amino acids a_{45}, a_{46}, a_{47}, thus $S_2=a_{45}+a_{46}+a_{47}$; E_3 has as active site the domain S_3 formed of the amino acids a_{70}, a_{71}, a_{72}, thus $S_3=a_{70}+a_{71}+a_{72}$; E_4 has as active site the domain S_4 formed of the amino acids a_{101}, a_{102} thus $S_4=a_{101}+a_{102}$.

When all the active sites function normally, the transformation of substance A into X takes place. There may be situations when those sites are inactive, so the enzyme can no longer intervene in that metabolic cycle. Inactivation of enzymes can occur as a result of mutations in the DNA molecule. Mutations whose effect is to change the structure of the gene corresponding to the enzyme (the causes of mutations can be chemical or physical [16]).

So, if outside the tissue made up of cells, which possesses the respective metabolic cycle, substance X is identified, then the situation is normal. When substances D, C, B, A are identified (there are inactivated the enzymes E_4, E_3, E_2, E_1), situation shown in Tab.1.

Table 1

The identified substance	Inactivated enzyme
D	E_4
C	E_3
B	E_2
A	E_1

Inactivation of enzymes is due to changes in nucleic acids, including changes in the position of various amino acids in the structure of the active site of the enzyme. For instance, if enzyme E_4 is inactivated, then substance X can no longer be synthesised (instead of it, substance D will be identified, and the metabolic cycle is no longer complete). The causes for the inactivation of enzyme E_4 are due to genetic mutations that affect the gene G_4, gene that allows the synthesis of the enzyme (the result of mutations being the modification of the amino acid a_{101} or a_{102} or both amino acids simultaneously). By this mechanism, the active site S_4 turns into inactive site, and this situation may also be found for the other enzymes E_3, E_2, E_1 (Tab.2).

Table 2

Enzyme	Amino acids that can suffer changes
E_4	a_{101}, a_{102}
E_3	a_{70}, a_{71}, a_{71}, a_{73}, a_{74}
E_2	a_{45}, a_{46}, a_{47}
E_1	a_{15}, a_{16}, a_{17}

At the same time, due to the inactivation of different enzymes, modified metabolic cycles occur (Fig.5).

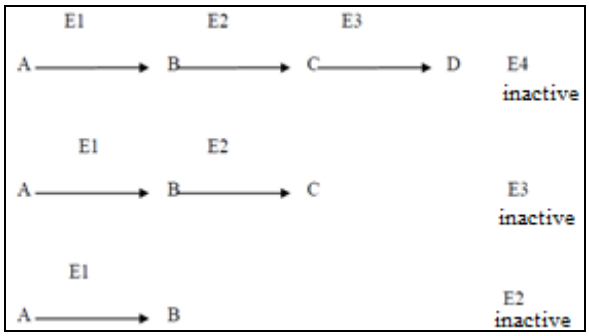

Figure 5. Modified metabolic cycles. *Legend of the picture: E_4 inactive / E_3 inactive/ E_2 inactive.*

The change in the structure of an enzyme is due to the change in a gene in the DNA molecule.

Analysing in detail, the following can be found: the enzyme E_4 is inactivated, so its active site (S_4) is inactivated (site consisting of two amino acids a_{101} anda_{102}). For the two amino acids, there are correspondent two codons c_{101}, c_{102}, respectively, in the structure of the gene G_4, with the following structure c_{101}=GAG and c_{102}=ATA. Due to one mutation in c_{101}, instead of GAG, we will find AAG, so the first nitrogenous base G has been replaced by the nitrogenous base A, and the whole site S_4 is inactivated. The S_4 site will transform and have the structure AAG+ATA, turning inactive (Tab.3).

Table 3

Amino acid	Codon	Codon structure
a_{101}	c_{101}	GAG
a_{102}	c_{102}	ATA

For the enzyme E_3, with the active site S_3, and having the structure $S_3=a_{70}+a_{71}+a_{72}+a_{73}$, there are correspondent to the amino acids the codons c_{70}, c_{71}, c_{72}, c_{73}, c_{74}, having the related structures presented in Tab.4.

Table 4

Amino acid	Codon	Codon structure
a_{70}	c_{70}	TTT
a_{71}	c_{71}	ATA
a_{72}	c_{72}	TAT
a_{73}	c_{73}	GAG

If a nucleotide is modified, for example, at a_{72}, instead of TAT we have TGT, the codonc_{72} will encode another amino acid, and the site S_3 will be inactivated (just like the enzyme E_3). For the enzyme E_2, the active site S_2 has the following structure $S_2=a_{45}+a_{46}+a_{47}$ (Tab.5).

Table 5

Amino acid	Codon	Codon structure
a_{45}	c_{45}	AAG
a_{46}	c_{46}	GAA
a_{47}	c_{47}	ATT

A codon change c_{47} from TTA to TGA and a_{47} will be replaced with a*47, which leads to the modification of the structure of the S_2 site, its inactivation, and, implicitly, of the E_2 enzyme. In the case of the E_1 enzyme, the S_1 site has the following structure $S_1=a_{15}+a_{16}+a_{17}$ (Tab.6).

Table 6

Amino acid	Codon	Codon structure
a_{15}	c_{15}	TGG
a_{16}	c_{16}	GAT
a_{17}	c_{17}	AAT

A change in the structure of a_{16} from TGA to TGG will result in another a_{16} amino acid, which may block the S_1 site, and thus the E1 enzyme. In general, the situation can be remedied if the initial chemical species is identified (A in the analysed case and the inactivated enzyme is E_1), by introducing it the metabolic cycle can be restored. (Fig.6).

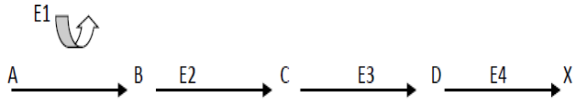

Figure 6. Explicative on the restoring of the metabolic cycle.

The situation is the same for the other enzymes (E_2, E_3, E_4), and the various metabolic cycles that may occur are shown in Fig.7. More complex cycles can occur (Fig.8). As can be seen, two enzymes E_4 and E_5 act on the intermediate chemical species D, each eventually leading to the appearance of two different chemical species, denoted by X and Y.

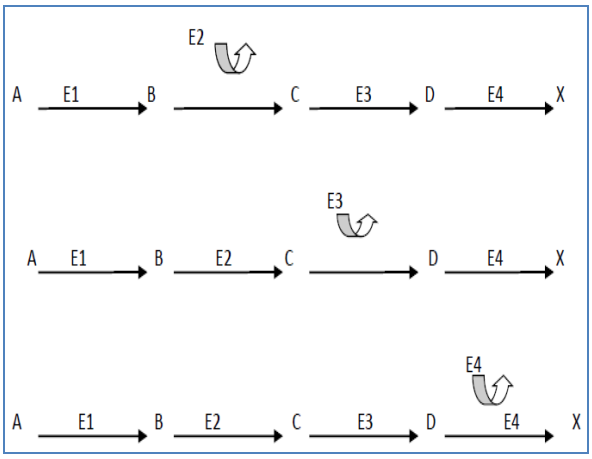

Figure 7. Simplified representation of a metabolic cycle.

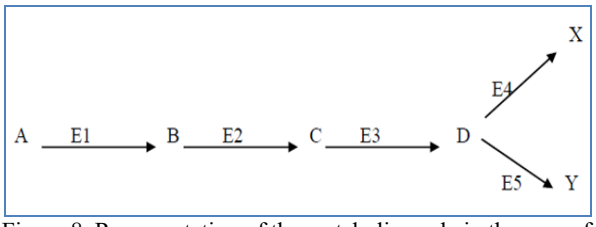

Figure 8. Representation of the metabolic cycle in the case of the intermediate chemical species D.

A similar situation can occur in the case of the intermediate chemical species C, on which two or more enzymes (E_3 and E_4) can act, leading to the appearance of substance D and the continuation of the metabolic cycle (in the case of E_3), or of the final substance. X (in the case of E_4 - Fig.9).

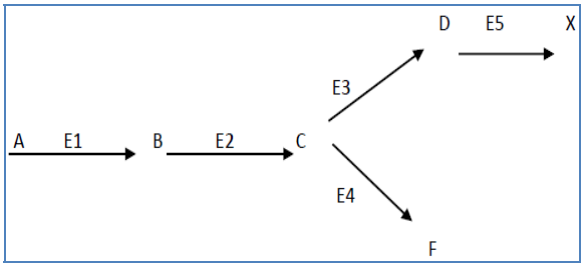

Figure 9. Representation of the metabolic cycle in the case of the intermediate chemical species C.

For the intermediate chemical species B, we may have the situation given in Fig.10.

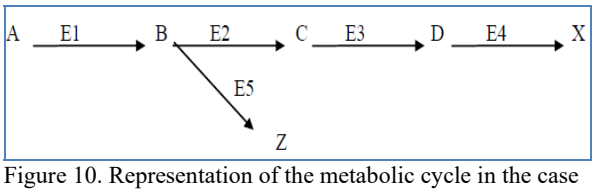

Figure 10. Representation of the metabolic cycle in the case of the intermediate chemical species B.

We can identify a tree structure of a metabolic cycle (Fig.11). The situation presented is often encountered in the functioning of a cell, where various enzymes may be in competition for the substrate. And in this case, as in the case of a linear metabolic cycle, the various mutations that occur at the codons encoding amino acids can lead to the interruption of the metabolic cycle or to change its structure (in this case the final products being very different from the products to be obtained normally).

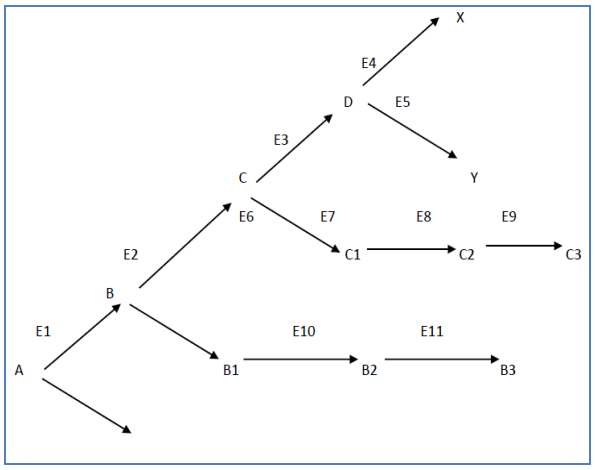

Figure 11. Ramified structure of a metabolic cycle.

A challenge of the last decade is the analysis of the variation of the structure of a gene over time. In this case, in addition to the codons that make up the gene, a variable is introduced in the description of the structure of the gene, so we will have the equation:

$$S_g = f(c_1....c_n, t), \quad (5)$$

where c_1, ... , c_n are the codons forming the gene (n= finite), and t represents time.

The purpose is to analyse is the situation when at time t_1 we have a structure $S_1 = f(c_1, ..., c_n, t_1)$, and at time t_2 we have a slightly modified structure of the same gene $S_2 = f(c_1, ... , c_n, t_2)$. This process occurs naturally and represents the aging of an organism. A gene is an entity with an extremely dynamic structure that can be a new field of application.

Genetic studies have traced the path of the virus from animal hosts to humans, tracking genetic variations in the protein S (spike).

In the case of coronaviruses, structural proteins are encoded, including the glycoproteins of the envelope (S), the envelope (E), the membrane (M) and nucleocapsid (N). In addition to the genes encoding structural proteins, there are accessory genes that are species-specific and dispensable for virus replication.

3. Employed recursive algorithms

One of the recent challenges is the computing connection of the analysis of a system with that of the functioning of a metabolic cycle. From the genetic analysis of a body (the karyotype of the body) a situation similar to a "if computer" can be highlighted. For example, if the human chromosome 5 is normal, the individual is considered to be biologically viable, and if a deletion of the short arm of the chromosome is found, a disease called "cri du chat" will occur (manifested by microcephaly, mental retardation and various heart conditions). The following are sequences from the program that transposes the described

analysis (Fig.12) and that uses recursive backtracking [17].

```
s// initializes the codon lists, reading //
encoding, from the codon file.txt
intinitializationcodons()
{ FILE*in;int i;
if((in=fopen("codons.txt","rt"))==NULL)
 {fprintf(stderr,"Do not open CODONI.TXT\n");
return 1;}
i=0; while(!feof(in))
  { fscanf(in,"%s%s", c[i],d[i]);i++;}
fclose(in);printf("Identify codons\n");
for(i=0;i<64;i++)
  {printf("%s%s\n",c[i],d[i]);}
return 0;}
//detects a codon index
intdetectcodon(char*s)
{ int i; for(i=0;i<64;i++)
if (strcmp(s,c[i])==0)return i;
return -1;}
// reading an issue from the issue file.txt
intreadingissue ()
{ FILE*in;int i;
if((in=fopen("probl.txt", "rt")) == NULL)
  {fprintf(stderr, "Do not open ISSUE.TXT\n");
return 1;}
n=0; while(!feof(in))
  { fscanf(in,"%s",t[n]);n++;}
fclose(in);n--;
printf("Reading issue (X= absent codon):\n");
for (i=0;i<n;i++)
printf("%s ",t[i]); printf("\nMeaning:\n");
for (i=0;i<n;i++)
  { x[i]=detectcodon(t[i]);
if(x[i]==-1)printf("Codon_lipsa");
elseprintf("%s",d[x[i]]); }
printf("\n"); return 0;}
//detects the dimension of the issue, number of
"X"
intdimensionoftheissue()
{ int i,m=0; for(i=0;i<n;i++)
if(x[i]==-1)m++;return m;}
// processes the solution found by the principle
of
// the meter, ie displays it coded
voidprocess_solution()
{ int i,j;printf("Found solution:\n");j=0;
for (i=0;i<n;i++)
  { if(x[i]==-1){printf("X=%s",c[y[j]]);j++;}
elseprintf("%s",c[x[i]]); }
printf("\nNamely:\n");j=0;
for (i=0;i<n;i++)
  { if(x[i]==-1)
    { printf("X=%s",d[y[j]]);j++;}
elseprintf("%s",d[x[i]]);}
printf("\n"); }
// backtracking, recursive implementation,
// for the counter principle
intbacktrackingrecursive (intlevel)
{ if(level>100)return 1;
if(level==m+1)process solution();
elsefor(int i=0;i<64;i++)
  { y[level]=i;
if(backtrackingrecursive(level+1))
return 1;}
return 0;}
//solving the issue by the "counter method"
intsolvingissue()
{ int m; m=dimensiuneaproblemei();
printf("Solutionam o problema%d\n",m);
backtrackingrecursive(0);
return 0;}
intmain()
{ if(initialisationcodons())return 1;
if(readingissue()) return 1;
if(solvingissue())return 1;
return 0;}
```

Figure 12. Software implementation of the algorithm.

4. Conclusions

Because the issue addressed is very broad, complex and extremely dynamic, in this paper only a few aspects were highlighted regarding the major importance of a protein for the proper functioning of a living body. As shown, the modification of a single parameter that characterizes a protein structure (the length of the protein chain), can lead to an undesirable effect and more precisely, to the appearance of biologically unviable organisms. By means of biochemical methods, cumulated with recursive algorithms implemented in high-level languages, the various chemical species can be identified, which appear as a result of the structural transformations of proteins. If the structure of a protein and its mode of action are known, various drugs can be prepared to prevent and fight a possible adverse effect of that protein. Obviously the most acceptable procedure is a mathematical description of the process, considering as many restrictive conditions as possible, followed by a software application as possible objective. The problem, in reality, is much more difficult because all the factors involved in some biological processes are not yet known. In the case of metabolic cycles, the knowledge of the structure of functional proteins and especially the dynamics of their structure allow the highlighting of various variants of the respective cycle. This way, by knowing the various "hot spots" in the structure of a gene (a gene that allows the synthesis of a functional protein, participating in the dynamics of the cycle) we can create various applications that allow a more accurate interpretation of the cycle.
Also, mathematical analysis of the structure of coronavirus tips can be used to discover vaccines needed by the human body to produce antibodies. These tips that help the virus "bind" and infect healthy cells are also those that allow the immune system to identify the virus.

References

[1,10,11,12,14]. Voet, D., Voet, J.G., *Biochemistry. 3*., Ed. Auflage. John Wiley & Sons Inc., London, 2004.
[2,13]. Andersen, K.G., Rambaut, A., Lipkin, W.I.. et al., *The Proximal Origin* of SARS-CoV-2, Nature, 2020.
[3]. Cui, J., Li, F., Shi, Z.L. *Origin and Evolution of Pathogenic Coronaviruses*. Nature Reviews: Microbiology, 2019.
[4]. Lewin, B., *Genes*, P., Ed. Prentice Hall, 2004.
[5,7]. Lesk A., *Introduction to Bioinformatics*, Oxford University Press Inc., New York, 2002.
[6] Bruice, P., *Organic Chemistry*, Prentice Hall, 2012.
[8,16]. Lupea, A.X., *Biochimie fundamente*, (*Basic of Biochemistry)* Academy Publishing House, Bucharest, 2007.
[9]. Benvenuto, D., Giovannetti, M., Ciccozzi, A., Spoto, S., Angeletti, S., Ciccozzi, M.. *The 2019-New Coronavirus Epidemic: Evidence for Virus Evolution*. Journal Med Virol. 2020.
[15]. Ciobanu, M., Dărăban, A., Flowers, D., Iacob, A., Pleşu, M*., Probleme de genetică. Aplicaţii ale matematicii în genetică (Matters of Genetics. Applications of Mathematics in Genetics)*, Ed. InfoData, 2016.

[17]. Popescu M.C., Olaru O., *Simularea numerică a proceselor (Numeric Simulation of the Processes)*, The Printing House of the University of Tg. Jiu, 1996.

Ovidiu Costica UNGUREANU - PhD Lecturer at "Vasile Goldis" Western University of Arad, Romania licensed in agronomy and biology master's degree in Plant Protection, PhD in Agronomy. Scientific research: 90, books and university text-books: 7, research grants: 10. Research domains: Plant Physiology, Phytotechnics, Phytopathology, Entomology.

Marius-Constantin POPESCU (b. May 19, 1965), PhD in Electrical Engineering 2002, University of Craiova, Romania, MSc in Automation and Computers (1990) and in Informatics (1995), Professor at Faculty of Economics, Engineering and Computer Science, "Vasile Goldis" Western University of Arad, Romania. Author and co-author 25 books and more than 250 papers. Research interests involve signal process, modelling of simulation processes. Preoccupations in the above-mentioned fields are reflected in a number of over 25 projects, as manager or member, funded by national or international research authorities or county council.

Elena UNGUREANU - PhD Lecturer at University of Agricultural Sciences and Veterinary Medicine Iasi, Romania, licensed in chemistry and physical (2004), master's degree in the paper Science and Engineering (2005), PhD in Chemical Engineering: (2008). Scientific research: 107, B\books and university text-books: 13, research grants: 6. Research domains: Chemical synthesis and analysis of organic and inorganic compound's structure; Environmental Analysis; Studies of different disperse systems and of transport phenomena;

Calin-Gabriel SARLA (b. Sep. 13. 1963), (2008) PhD in Physical Education and Sport, University of Pitesti, Romania, (2000) M.Sc and BSc in Physical Education and Sport, University of Craiova, Romania, (1991) BEng. in Technological Electro Mechanics, University of Oradea, Romania. Author of 5 books, editor coordinator 3 books and (co-) authored 25 papers to international high level. Associate Professor, at "Vasile Goldis" Western University of Arad, Romania, Department of Physical Education and Sport, Arad, Romania Field of interest: The orchestration of specific real processes through the transformation of statistic indices in instructional objectives of the training processes. Large understandings about how it works together different components of human body.