

A Quantum Approach to the DNA Functioning

A. Nicolaidis

Department of Theoretical Physics
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece

Abstract

In the present work we prime the notion that DNA is an information processing system, receiving – registering – transferring information. In the pursuit of an inherent logic in DNA functioning, we explore the possibility that quantum logic might serve this purpose. We use the quantum formalism to describe the DNA dynamics and as a byproduct we obtain the DNA vacuum. The DNA vacuum, in clear analogy to the quantum vacuum, is a collection of virtual DNA bases. An essential aspect of the DNA functioning is the complementarity relation R , which binds the pairs A-T, G-C, and generates the replication process. Further in an effort to codify DNA, we introduce Gödel's numbering for a DNA strand, assigning a specific natural number to each individual strand. This numbering allows a quantitative measure of the difference among the various DNA strands. Considering also that the four DNA bases constitute an “alphabet”, we may assume the task to examine if DNA is a “language”.

Introduction

Considering the entire evolution, from cosmology, through biology, to language, we gather that biology verges on natural sciences and linguistics. Natural sciences (physics, chemistry) prime the notion of structure. There are elementary constituents, the “atoms”. The “atoms” through interactions, get composed to give larger structures (from quarks, to protons and neutrons, nuclei, atoms, molecules, stars, galaxies). Linguistics primes the notion of information. It introduces the “sign”, or the “word”, which denotes and refers to another object.

We expect that biology shares features from both forms of knowledge, natural sciences and linguistics. Indeed in biology we encounter the biological “atoms”, the four nucleotide molecules (adenine, guanine, cytosine, thymine). Further the four nucleotides get composed to form larger structures, the DNA sequences, amino acids, proteins. From another point of view the four nucleotides may be considered as not simply the constituents of biological structures, but as the “letters” of a language. These “letters” give rise to biological “words”, “phrases”, “sentences”. The biological “words” or “phrases” act like signs, receiving – registering – transferring information, executing specific functions, favoring or disfavoring a biological process. It is an open and a highly interesting question if the biological “text” follows an internal logic, or a syntax.

In the present work, in the search of a biological syntax, we would like to explore the possibility that quantum logic might be related to the internal logical functioning of DNA. Most of the scientific edifice relies on Aristotle’s logic (with the law of excluded third: it is either A or its opposite $\neg A$). Quantum Mechanics defies though common sense and common logic.

In ref. [1, 2] we have suggested that the relational logic of C. S. Peirce [3, 4] may serve as the conceptual foundation of quantum mechanics and string theory. Within relational logic, relation is the primary irreducible datum and everything is expressed in terms of relations. We are led to reorient our thinking and consider that things have no meaning in themselves, and that only the correlations between them are “real”. Few examples of relations R_{ij} might be indicative: the transition from a state j to a state i , the proof of a theorem i starting from a theorem j , the transformation of a metabolite j to a metabolite i . Relations may be composed and a third transitive relation emerges following the rule

$$R_{ij}R_{kl} = \delta_{jk}R_{il} \quad (1)$$

In quantum theory the quantum states are living in a Hilbert space. Transitions between the quantum states are realized through relations or projection operators. We propose then to represent the four nucleotides (A, G, C, T) as states living in an abstract space. The pairing A-T and G-C is achieved through a corresponding projection operator. Furthermore we introduce the DNA vacuum. In a similar way that the quantum vacuum is not empty, but is a collection of virtual particles, the DNA vacuum is a collection of all possible bases. The notion of the DNA vacuum will appear very useful in better understanding the replication process. In the next section we introduce the quantum formalism and we use it in order to represent and better understand the DNA functioning. In the third section we consider a DNA strand as a “theorem” in logic. Following Gödel’s numbering of mathematical theorems [5], we represent the DNA strand by a unique natural number, a product of prime numbers. This encoding of DNA

may allow a quantitative measure of the difference among the various DNA strands. Also it would help for a faster and more efficient analysis of DNA through data mining techniques. In the last section we present our conclusions and indicate directions for future research.

Quantum formalism for DNA

We suggest that the quantum formalism is apt in order to describe the inherent DNA dynamics. Let us remind the essential of the quantum formalism, using the Dirac bra-ket notation. A quantum ket state $|S_i\rangle$ is living in an abstract Hilbert vector space. Next to the Hilbert space there is a dual Hilbert space where live the bra states $\langle S_i|$. A bra state is the transpose and conjugate of the corresponding ket state. We can define then two products. The “inner product”

$$A_{ij} = \langle S_i | S_j \rangle \quad (2)$$

is a number and indicates the affinity or similitude between the $|S_i\rangle$ and $|S_j\rangle$ states. Since the “inner product” is in general a complex number, a real measure of the affinity is obtained through the probability p ($0 \leq p \leq 1$)

$$p = A_{ij} A_{ji} \quad (3)$$

Like in all vector spaces, in our Hilbert vector space, there are “base vectors” which are orthonormal, satisfying

$$\langle n | m \rangle = \delta_{nm} \quad (4)$$

We define also the “outer product”

$$R_{ij} = |S_i\rangle\langle S_j| \quad (5)$$

R_{ij} stands for a relation, which expresses the transition from the initial state $|S_j\rangle$ to the final state $|S_i\rangle$. It can be considered also as a projection operator which allows as incoming state $|S_j\rangle$ and as outgoing state $|S_i\rangle$. Clearly R_{ij} , thus defined, satisfies the composition rule equ. (1).

The hereditary information of an organism is encoded in the DNA. The DNA is a macromolecule composed of two polynucleotide chains with a double-helical structure. There are four distinct bases, building elements of the genetic information: adenine, cytosine, guanine, thymine, abbreviated A, C, G, T respectively. A single strand DNA is simply a chain of nucleotides where two consecutive nucleotides are bound together by a strong covalent bond. Each single strand has a natural orientation. This orientation is due to the fact that one end of the single strand has a free 5' phosphate group and the other has a free 3' deoxyribose hydroxyl group. The most important feature of DNA is the Watson-Crick complementarity of bases. Bonding between single strands occurs by the pairwise attraction of bases: A bonds with T and G bonds with C. The pairs (A, T) and (G, C) are therefore known as complementary base pairs. The classical double helix of DNA is formed when two separate strands bond. Two requirements must be met for this to occur; firstly, the strands must be complementary, and secondly, they must have opposite orientations.

We may now borrow quantum ideas and apply them to the DNA functioning. We suggest that the four bases A, C, G, T, are states belonging to a Hilbert space and we represent them by $|B_1\rangle, |B_2\rangle, |B_3\rangle, |B_4\rangle$

respectively (clearly the number correspondence is arbitrary). The states satisfy the orthonormal condition

$$\langle B_i | B_j \rangle = \delta_{ij} \quad (6)$$

The bonding between the bases $|B_1\rangle$ and $|B_4\rangle$ is achieved through a relation (or equivalently a projection operator) P

$$P = |B_1\rangle\langle B_4| + |B_4\rangle\langle B_1| \quad (7)$$

The bonding between the bases $|B_2\rangle$ and $|B_3\rangle$ is achieved in a similar fashion through a relation Q

$$Q = |B_2\rangle\langle B_3| + |B_3\rangle\langle B_2| \quad (8)$$

The full complementarity is realized with the relation

$$R = P + Q \quad (9)$$

Notice that

$$R|B_i\rangle = |B_j\rangle \quad (10)$$

where $|B_i\rangle$ and $|B_j\rangle$ are complementary pairs. R acts like a “mirror” operation with A, T being the images of each other (similarly for the C and G bases). R^2 is an idempotent operator, i.e it is the unity operator $\mathbf{1}$

$$R^2 = \sum_i |B_i\rangle\langle B_i| = \mathbf{1} \quad (11)$$

and

$$R^2|B_j\rangle = |B_j\rangle \quad (12)$$

Let us concentrate on the 2-state problem specified by the P relation (similarly for the Q relation). $|B_1\rangle$ and $|B_4\rangle$ are represented by the columns

$$|B_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, |B_4\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (13)$$

and P relation is represented by the matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (14)$$

The eigenstates of P are the states $|B_+\rangle$ and $|B_-\rangle$

$$|B_+\rangle = \frac{1}{\sqrt{2}}[|B_1\rangle + |B_4\rangle], |B_-\rangle = \frac{1}{\sqrt{2}}[|B_1\rangle - |B_4\rangle] \quad (16)$$

Notice that the states $|B_+\rangle$ and $|B_-\rangle$ are transformed to each other through the relation S

$$S|B_\pm\rangle = |B_\mp\rangle \quad (17)$$

with

$$S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (18)$$

The relations P and S anticommute

$$PS + SP = 0 \quad (19)$$

We obtain a better understanding of the relational dynamics by using a double line representation of the general relation R_{ij} . Each distinct state or base $|B_i\rangle$ is represented by a specific line, with a downward (upward) arrow

attached to the initial (final) state. In this sense we picture $R_{14} = |B_1\rangle\langle B_4|$ by

$$\begin{array}{ccc}
 & \uparrow & \\
 A & | & \\
 & \downarrow & \\
 & T &
 \end{array} \quad (20)$$

The relation $R_{11} = |B_1\rangle\langle B_1|$ is pictured by

$$\begin{array}{ccc}
 & \uparrow & \\
 A & | & \\
 & \downarrow & \\
 A & &
 \end{array} = \text{bubble with arrow} \quad (21)$$

The identity, eq. 11, appears as a collection of bubbles

$$\begin{array}{cccc}
 \text{bubble with arrow} & + & \text{bubble with arrow} & + & \text{bubble with arrow} & + & \text{bubble with arrow} \\
 A & & C & & G & & T
 \end{array} \quad (22)$$

We consider that the identity expression and its picture represent the DNA vacuum. The DNA vacuum is not empty but it is full of DNA bases, which are continuously annihilated and created. This restless DNA vacuum resembles the QCD vacuum. QCD (Quantum Chromodynamics) describes the strong interactions among quarks, the fundamental entities of hadrons

(proton, neutron, pion etc). The QCD vacuum is a sea of restless quarks and antiquarks.

We may move further and picture the composition of relations, equ (1). For example the composition

$$R_{14}R_{41} = R_{11} \quad (23)$$

will look like



Clearly the above diagram looks like a string diagram, but at the same time it represents the DNA replication process starting from the DNA vacuum.

Next to the discrete operation R , we may consider a continuous operation $\exp(\alpha R)$, where α is a parameter. A Taylor expansion provides

$$\exp[\alpha R] = \mathbf{1} \cos \alpha + R \sin \alpha \quad (25)$$

The operation $\exp(\alpha R)$, acts like a rotation. Consider the bonding $|T\rangle\langle A|$. Upon rotation it is transformed to

$$\exp(\alpha R)[|T\rangle\langle A|] \exp(\alpha R) = \cos^2 \alpha |T\rangle\langle A| + \sin^2 \alpha |A\rangle\langle T| + \frac{1}{2} \sin 2\alpha [|T\rangle\langle T| + |A\rangle\langle A|] \quad (26)$$

For $\alpha = \frac{\pi}{4}$ the final result looks like

$$\frac{1}{2} [|T\rangle\langle A| + |A\rangle\langle T| + |T\rangle\langle T| + |A\rangle\langle A|] \quad (27)$$

We observe that next to the initial bonding $|T\rangle\langle A|$, we obtain in a formal way the bonding $|A\rangle\langle T|$, plus a collection of “sea” DNA bases. Thus the replication process is a direct manifestation of the Watson-Crick complementarity principle.

Gödel’s numbering for DNA

A single DNA strand is a chain of bases, where two consecutive bases are bound together by a covalent bond. We may consider that each of the four bases stands for a letter, establishing a 4-letter alphabet. A succession of these letters in a DNA strand may represent a word or a phrase in a biological language. From another point of view each base may represent a symbol in an axiomatic system. A succession of these symbols may correspond to a mathematical theorem.

Along this line, we are entitled to be inspired by the work of Kurt Gödel [5]. Gödel made an immense impact upon scientific and philosophical thinking in the 20th century, by establishing the incompleteness theorem. To prove this theorem, Gödel developed a technique, now known as Gödel numbering, which codes formal expressions as natural numbers.

The starting point in Gödel’s numbering is to assign a positive natural number $\#(s)$ to each of the symbol s , in a fixed but arbitrary way. This is the Gödel number of the symbol. We adopt

$$\#(A) = 1, \#(C) = 2, \#(G) = 3, \#(T) = 4 \quad (28)$$

For a succession of DNA bases, for example the four bases AGCA, we pick up the first four prime numbers and we raise each of them to the corresponding Gödel number. Thus

$$w(AGCA) = 2^1 3^3 5^2 7^1 = 9450 \quad (29)$$

is a unique natural number, the Gödel's number w for the specific DNA strand. Inversely, given a Gödel's number, we can decode it and find the DNA strand it represents. For example consider the number $w = 3240$

We factorize it as a product of the prime numbers 2, 3, 5, 7, 11 and find how many 2, 3, 5, ... are hidden in the number. We find out that

$$w = 2^3 3^4 5^1 \quad (30)$$

and therefore the above w represents the DNA strand GTA

Gödel's numbering allows us to obtain a quantitative measure of the difference among the various DNA strands. Consider two strands represented by the Gödel's numbers

$$w_1 = 2^{y_1} 3^{y_2} 5^{y_3} 7^{y_4} \dots p_k^{y_k} \quad (31)$$

$$w_2 = 2^{q_1} 3^{q_2} 5^{q_3} 7^{q_4} \dots p_k^{q_k} \quad (32)$$

where 2, 3, 5, 7, ... p_k are the prime numbers and $y_j(q_j)$ represents the

j -th base in the first (second) strand. The difference D among the two strands is

$$D = \frac{w_1}{w_2} = 2^{(y_1/q_1)} 3^{(y_2/q_2)} 5^{(y_3/q_3)} \dots p_k^{(y_k/q_k)} \quad (33)$$

We notice though that the identification $\#(s)$ is arbitrary and therefore D cannot serve as an objective measure. Rather, considering for example w_1 as a reference strand, we define the difference Δ between the reference strand and another strand by

$$\Delta = \prod_j \frac{1}{p_j} \quad (34)$$

where p_j stand for all those prime numbers where the DNA bases differ in the corresponding j places. Imagine the evolution of DNA. We may assume that everything started with an “original” DNA, which upon differentiation provided the subsequent plethora of DNA. We expect that the differences are located at late positions, the corresponding p_j are large and correspondingly Δ is small. On the other hand if the differences are located at the initial places, then Δ is relatively large.

In a DNA “language” we encounter motifs, that is a precise combination of bases which serves a specific function. Assuming that a motif is a combination of n bases, starting at the position k , its Gödel’s number $M_{k,n}$ is the product of n prime numbers raised to the appropriate powers, representing the corresponding bases

$$M_{k,n} = p_k^{s_i} p_{k+1}^{s_j} \cdots p_{k+n}^l \quad (35)$$

Within a then a huge DNA collection we can search, through data mining techniques, for the existence of the different $M_{k,n}$.

It should be added, that in order to evaluate similarity or difference in discrete hierarchical biological systems, it has been proposed also to use the

ultrametric distance, notably in its p-adic version[6]. This approach has been useful in classifying the codons.

Conclusions and future research directions

There is long standing effort for a “quantum biology”. It involves how purely quantum effects, like entanglement and coherence, might be of relevance in biological systems [7 – 12]. In the present work we do not consider DNA as a quantum system per se. Rather we suggest that DNA’s inherent logic is the relational logic, the same logic that appears governing quantum mechanics and string theory. The emphasis is not in terms of structure, but in terms of information processing. The four DNA bases may be considered as the “letters” of a DNA alphabet and a DNA strand as a “word” or “sentence” in a biological text. The analogy with quantum logic prompted us to use the quantum formalism in order to describe the DNA dynamics. As a by-product of this approach we obtained the DNA vacuum, a collection of annihilated and created DNA bases. An important relation for the DNA is what we defined as the relation of complementarity R . Relation R guarantees the bonding A-T and G-C and generates the replication process. We have shown also that the relations introduced maybe represented by matrices and therefore the whole DNA functioning can be accounted for by a matrix mechanics. Within the spirit of matrix mechanics we can search for interactions among matrices distance apart, giving rise to a sort of “entanglement”.

The search of coding regions in a DNA sequence, encouraged us to use the Gödel's numbering in order to codify a DNA strand. This numbering allows to study the existence of a specific DNA "word" within a broader DNA sequence by a simple factorization process. A remarkable feature of languages is Zipf's law. This law dictates that the frequency f of each word in a text and its rank are related according to a power law. It would be most interesting to check if Zipf's law prevails in a biological text [13]. The possible correlation between two DNA strands can be approached also in a novel way. Denoting by x, y the Gödel's numbers of the strands, the implicit correlation may be expressed by a function $F(x, y)$, with arguments of the function the numerals x, y .

Acknowledgement

I would like to acknowledge useful discussions with Dr. Kosmas Kosmidis and his assistance in preparing the paper.

References

1. A. Nicolaidis, *Categorical Foundation of Quantum Mechanics and String Theory*, Int. J. Mod. Phys. A24: 1175 - 1183, 2009;
2. A. Nicolaidis (2013), *Relational Quantum Mechanics*, Advances in Quantum Mechanics, Prof. Paul Bracken (Ed.), InTech, DOI: 10.5772/54892.
3. C. S. Peirce, *Description of a notation for the logic of relatives, resulting from an amplification of the conceptions of Boole's calculus*

- of logic*, Memoirs of the American Academy of Sciences 9, pp 317-378 (1870)
4. C. S. Peirce, *On the algebra of logic*, American Journal of Mathematics 3, pp 15-57 (1880)
 5. Gödel, K., 1931, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik Physik, 38: 173–198. English translation in Gödel Collected Works I. Publications 1929–1936, S. Feferman et al. (eds.), Oxford: Oxford University Press. 1986, 144–195.
 6. B. Dragovich and A. Dragovich, *p-Adic Modelling of the Genome and the Genetic Code*, Computer Journal **53** (2010)
 7. M. Arndt, T. Juffmann and V. Vedral, *Quantum physics meets biology*, HFSP journal, vol. 3, 2009, 386-400
 8. N. Lambert, Y-N Chen, C-M Li, G-Y Chen and F. Nori, *Quantum biology*, Nature Physics, vol. 9, Jan. 2013
 9. G. Vattay, S. Kauffman and S. Niiranen, *Quantum biology on the edge of quantum chaos*, PLOS one, vol. 9, issue 3, e89017, March 2014
 10. I. Kominis, *The radical-pair mechanism as a paradigm for the emerging science of quantum biology*, Mod. Phys. Lett. B, **29**, 1530013 (2015)
 11. L. Kauffman, Self-reference, biologic and the structure of reproduction, arXiv:1512.04325
 12. D. V. Nanopoulos, *Theory of Brain Function, Quantum Mechanics and Superstrings*, invited talk at the “Physics without frontiers Four Seas Conference”, Trieste 1995, hep-ph-9505374

13. A. Tsonis, P. Kumar, J. B. Elsner, and P. A. Tsonis, *Is DNA a Language?*, J. Theor. Biol. (1997) **184**, 25-29