Many-sheeted DNA

M. Pitkänen,

November 30, 2016

Email: matpitka@luukku.com.
http://tgdtheory.com/public_html/.
Recent postal address: Karkinkatu 3 I 3, 00360, Karkkila, Finland.

Contents

Abstract

The problems of how genes code information about the morphology of organism and how this information is expressed, belong to the great puzzles of the developmental biology. A closely related mystery is the differentiation of cells. The notion of the genetic program is far from precise and it is not clear how close the analogy with a computer program is. There are also several problems which challenge the basic dogmas of genetics.

1. Only 1 per cent of DNA of human genome actually codes polypeptides. Eukaryote genes contain intron sequences which are transcribed into hnRNA but snipped of when hnRNA is transformed mRNA in process called slicing. The higher the evolutionary level of organism, the higher the fraction of introns is. Molecular Darwinists see introns as "junk DNA" but there is evidence that introns are far from junk. For instance, the splicing of intron contribution from hnRNA to give mRNA can give several different outcomes depending on the stage of development of the organism and introns are crucial for the effectiveness of immune system. Hence one can wonder whether intronic mRNA and protein mRNA could both form the real output of gene subprograms serving in some sense as input for other gene subprograms. This interpretation obviously conflicts with "gene-single protein" dogma in its basic form.
2. There are large amounts of highly repetitive DNA which is silent. One can wonder whether there is some fundamental mis-understanding involved. Could it be that this DNA is analogous to control DNA not transcribed to RNA and therefore not all useless. There is also active repetitive DNA.
3. There is large amount of silent DNA in control sections between genes. Could it be that this silent DNA expresses itself in some nonchemical manner? Chemical expression is very slow, translation rate being twenty aminoacids per second, and one can wonder whether life might have invented faster modes of gene expression and control of gene expression.
4. Plant genome is often by a factor of hundred longer than human genome. One could argue that the complexity of organism is measured by the length of the shortest program coding the organism. It is however not at all obvious how the genome of plants could be more redundant than human genome since repetitive sequences common to all animals are present. Introns are in fact more frequent in human genome. This suggests that some new unidentified degrees of freedom giving rise to complexity might be present and that the chemistry of DNA in the sense of standard physics is perhaps not all that is needed to understand genetic program.
5. Various self-organization processes such as self-assembly and de-assembly are very frequent in living systems. The problem how genes give rise to morphology of the organism is poorly understood. This forces to challenge the dogma of genetic determinism. One should be able to understand what is determined by genes and what is determined by self-organization and whether the genes of the standard physics are enough.

The reason why the above mentioned problems have turned out to be so untractable might be due to a wrong view about space-time. Many-sheeted space-time concept of TGD might be absolutely crucial for the expression of genetic code. Gene itself might be many-sheeted space-time structure coding faithfully the topology of the expression domain of gene. This many-sheeted structure of DNA could allow to understand the miraculous looking features of DNA replication and cell differentiation. TGD based view of evolution as p-adic evolution implied by the basic quantum theory, should be a crucial element of the picture. Together with p-adic length scale hypothesis, with Combinatorial Hierarchy model for genetic code allowing to interpret genes as Boolean statements, and general vision about quantum control and coordination based on a hierarchy of weakly coupled super conductors, the notion of many-sheeted DNA leads to precise quantitative predictions and a general model for genetic program. In particular, one can understand the mystery of introns. What interesting from the point of view of our consciousness is that it might be possible to interpret the Boolean statements represented by the exon and intron parts of genes as a physical representation for our belief system. Thus genes would code both matter– and mind like hardware of the living system.

The notion of magnetic body is central in the TGD inspired theory of living matter. Every system possesses magnetic body and there are strong reasons to believe that the magnetic body associated with human body is of order Earth size and that there could be hierarchy of these bodies with even much larger sizes. Therefore the question arises what distinguishes between the magnetic bodies of Earth and human body.

The vision about dark matter hierarchy labelled partially by a hierarchy of values of effective Planck constant coming as integer multiples of the ordinary Planck constant leads to a rather concrete view about the hierarchy of magnetic bodies and implies a natural generalization leading to the notion of super- and hyper genes. The original model assumption $\hbar_{eff} = \lambda^{k_d}\hbar_0$, $\lambda \simeq 2^{11}$, is however un-necessarily strong.

Super genes consist of genes in different cell nuclei arranged to threads along magnetic flux sheets like text lines on the page of book whereas hyper genes traverse through genomes of different organisms. Super and hyper genes provide an enormous representative capacity and together with the dark matter hierarchy allows to resolve the paradox created by the observation that human genome does not differ appreciably in size from that of wheat.

1 Introduction

The problems of how genes code information about the morphology of organism and how this information is expressed, belong to the great puzzles of developmental biology. A closely related mystery is the differentiation of cells. The notion of genetic program is far from precise and it is not clear how close the analogy with a computer program is. There are also several problems which challenge the basic dogmas of genetics.

1. Only 1 per cent of DNA of human genome actually codes polypeptides. Eukaryote genes contain intron sequences which are transcribed into hnRNA but snipped off when hnRNA is transformed mRNA in a process called slicing. The higher the evolutionary level of organism, the higher the fraction of introns is. Molecular Darwinists see introns as "junk DNA" but there is evidence that introns are far from junk. For instance, the splicing of the intron contribution from hnRNA to give mRNA can give several different outcomes depending on the stage of the development of the organism and introns are crucial for the effectiveness of the immune system [I3]. Hence one can wonder whether intron mRNA and exon mRNA could both form the real output of gene subprograms serving in some sense as input for other gene subprograms. This interpretation obviously conflicts with "gene-single protein" dogma in its basic form.

2. There are large amounts of highly repetitive DNA which is silent. One can wonder whether there is some fundamental mis-understanding involved. Could it be that this DNA is analogous to control DNA not transcribed to RNA and therefore not at all useless. There is also active repetitive DNA.

3. There is large amount of silent DNA in control sections between genes. Could it be that this silent DNA expresses itself in some non-chemical manner? Chemical expression is very slow, translation rate being twenty amino-acids per second, and one can wonder whether life might have invented faster modes of gene expression and control of gene expression. Also the question whether there is a relation to the typical frequency scales of brain consciousness of order 10 Hz, which can be related to the magnetic and Z^0 magnetic transition frequencies, can be raised.

4. Plant genome is often by a factor of hundred longer than human genome. One could argue that the complexity of organism is measured by the length of the shortest program coding the organism. It is however not at all obvious how the genome of plants could be more redundant than human genome since repetitive sequences common to all animals are present. Introns are actually more frequent in human genome. This suggests that some new unidentified degrees of freedom giving rise to complexity might be present and that the chemistry of DNA in the sense of standard physics is perhaps not all that is needed to understand genetic program.

5. Various self-organization process such as self-assembly and de-assembly are very frequent in living systems. The problem how genes give rise to morphology of the organism is poorly understood. This forces to challenge the dogma of genetic determinism. One should be able to understand what is determined by genes what is determined by self-organization and whether the genes of the standard physics are enough.

The reason why the above mentioned problems have turned out to be so intractable might be due to a wrong view about space-time. Many-sheeted space-time concept of TGD might be absolutely crucial for the expression of genetic code. DNA itself might involve many-sheeted space-time structures coding faithfully the topology of the body parts. This many-sheeted structure of DNA could allow to understand the miraculous looking features of DNA replication and differentiation of cells. TGD based view of evolution as p-adic evolution implied by the basic quantum theory, should be a crucial element of the picture. Together with the p-adic length scale hypothesis it leads to precise quantitative predictions and a general model for genetic program based on the many-sheeted space-time concept. The model explains also why introns are present only in eukaryotic genome. Most importantly, it seems that the statements represented by the dynamical intron-exon decompositions of genes and defining Boolean algebra, could represent our conscious beliefs and thus affect our behavior as conscious beings. Notice the beautiful connection between matter and mind: genes code the information, not only about the material structure of organism, but also about its belief system. Thus without introns, the pariah class in the society of bio-molecules regarded as "junk DNA" by always-so-imaginative reductionistic materialists, we would have no world views and belief systems! In this chapter TGD based view about genetic code and its realization are discussed in detail.

1.1 Many-Sheeted DNA

The replacement of the DNA of standard physics with many-sheeted DNA suggest surprisingly simple model for how organism's morphology is coded and decoded to DNA.

1. How the morphology of body is coded?
   The most striking feature of DNA is its one-dimensionality. According to work of Mae-Wan Ho, living systems are liquid crystals [D1]. Liquid crystals are effectively one-dimensional since the layers of the liquid crystal consist of homogenous liquid phase determined by macroscopic characteristics such as pH, temperature, ionic concentrations and electric fields. This suggests that the structural information coded into DNA could be essentially information about the macro- properties of the layers of liquid crystal. This would make 1-dimensional coding of the body plan using DNA sequences very natural. Kind of contraction of the body parts to DNA sequences having many-sheeted structure could be in question! This coding would preserve the topological structure of the many-sheeted space-time surface representing the expression domain of the gene. The structure of the expression domains of maternal genes and Hox genes [I11] controlling morphogenesis supports this picture.

2. How DNA is expressed?
   The very naive first guess is that during growth various thin space-time sheets associated with DNA gradually grow and are glued together by the join along boundaries contacts and form the space-time sheets associated with their expression domains. Somewhat exaggerating, many-sheeted DNA would represent only a particular developmental period of organism in which it is contracted to a thin thread. For instance, the cells determined to develop into eye are glued to the space-time sheet representing future eye and replication products belong also to this space-time sheet. Clearly, the gluing to the space-time sheet of the future expression domain would generate the needed long range correlation between cells in the expression domain. It must be emphasized that self-organization should play key role in this process: for instance, liquid crystal nature of the living matter should determine morphology to a high extent.

3. What makes differentiation and control of the morphogenesis possible?
   Differentiation must be explainable as a selective activation of transcription and although a local process, involves also top-down control making possible a precise timing. Concentration gradients for the transcription factors, that is proteins controlling transcription, are certainly crucial in this respect. When the concentration of the protein falls below a critical value, the truth value of a statement representing input to some gene program modules changes. This leads obviously to spatial patterns of gene expression resulting from branching of gene programs. For instance, the development of organs should result as a combination of genetic control of this kind plus self-organization. Join along boundaries bonds between

gene space-time sheets and larger space-time sheets or genes and control regions of chromosome make possible quantum control of genetic expression based on phase gradients of the super conducting order parameters and resonant Josephson frequencies which correspond to magnetic transition frequencies of genes, control regions or related substructures. It could also be that the # contacts (wormhole contacts) from genes to various space-time sheets representing body parts provide the interaction with the classical fields of the macroscopic space-time sheets representing body arts and controlling the activity of a particular gene. In any case, the fact is that the action mechanisms of transcription factor proteins in eukaryotes are not understood. The mechanism is not purely chemical one since transcription factors are often located quite far from the promoter region. Electromagnetic oscillations with resonant frequencies could be in question. In absence/presence of oscillation gene is activated.

1.2 Realization Of The Genetic Program

TGD suggests concrete ideas about how organism can act as a conscious computer [**?**] Genome represents possible statements consistent with a fixed atomic statement (single element set in set-theoretic representation of Boolean algebra) represented as sequences of DNA. DNA triplets define basic axioms of the axiom system in question. Active genes which have coded some minimum about of intronic mRNA and protein coded by the exon part of the gene give rise to conscious experience about the truth of the statement represented by the gene. Otherwise the truth value of this statement is ill-defined and not consciously experienced. Truth values of the statements representing conclusion of the statement represented by gene in turn act as premises for the statements represented by some genes and these genes in turn activate and give rise to experience of the truthness of corresponding statements. In this manner genetic program proceeds and gives rise to a sequence of experiences about truthness of statements represented by the genome. Note however that the experiences of truthness at DNA level need not correspond to our conscious experiences: entire hierarchy of selves having connection to DNA level is involved.

The beautiful feature of this realization of genetic program is that no cables for signal transmission are needed. The genetic program is also extremely robust and flexible unlike ordinary symbol based programs in which the change of the value of single binary digit can lead to a catastrophe. Furthermore, spatial patterns of gene expression develop naturally: gene in give cell producing transcription factor affects only finite region of space since subcritical concentration of the transcription factor means effectively its total absence. This can lead to intricate structural patterns of gene expression and determination of cells making possible differentiation. The translation of average protein requires 20-60 seconds and the cognitive processes of ours which possibly occur at DNA level must be rather slow. Time scale of emotions is however slow as is also higher level abstract thinking.

Both introns and exons represent statements which are true if the premises of the gene statement are true. Simple model for how introns can be separated from the exon part of the genetic module explains the many mysterious properties of introns elegantly and introns become an absolutely essential element of the genetic program. In particular, addition or subtraction of a marking - "comment sign" - to nucleotide changes the nucleotide from exonic to intronic or vice versa. Thus this marking serves essentially as a binary digit telling whether nucleotide belongs to exon or intron. Unless physical realization of these markings poses any additional constraints, comments signs can be dropped anywhere in gene and this means that same gene can be expressed in 2^N manners, where N denotes the number of basic units in the maximal decomposition into exons and introns. Obviously, an interpretation as a representation for the statements of Boolean algebra for statements about N basic statements suggests itself strongly: perhaps each eukaryotic gene represents Boolean algebra! If also intron-exon decomposition is assumed to be dynamical then the number of exon-intron decompositions in gene consisting of N DNA triplets is

$$M = \sum_{k=1}^{N} \left(\begin{array}{c} 3N \\ 3k \end{array} \right) \quad .$$

The premises of the gene statements are represented in the operator sequences associated with the gene. Intron $\leftrightarrow$ exon transformation induced by the addition or cancellation of "comment signs" associated with nucleotides leads to a generalization of the operon model for the regulation

of gene transcription. The protein coded by introns which have become an exon part of gene serves as a repressor of the gene expression for original exons. The shifting back to a mode in which exons are coded to protein means that the coding of the supressor protein stops and the genes whose activation depends on the output of the gene in question are automatically activated.

A cautious proposal is that many-sheeted genes could represent a hierarchy of conscious beliefs: genome would be a collection of Boolean algebras represented by genes. In case of prokaryotes these Boolean algebras contain only one expressed element; in case of eukaryotes number of elements can be much larger but again totally intronic gene is not expressed. Maybe DNA could code thoughts and proteins emotions associated with them.

1.3 Are Non-Chemical Transcription Factors And Non-Chemical Gene Expression Possible?

Enhancers and silencers affect gene expression in a non-local manner difficult to understand purely chemically. There are also tissue specific transcription factors. The notion of many-sheeted DNA suggests the possibility of non-chemical transcription factors. Classical em and Z^0 fields are especially interesting possibility as far as gene expression and its control in long length scales are considered. The general quantum control and coordination mechanisms based on Josephson currents flowing between gene space-time sheets and larger space-time sheets and making possible comparison circuits, clocks, alarm clocks and novelty detectors, suggest themselves. Indeed, the frequency scale 20 amino-acids per second for the translation process corresponds to a typical scale for the magnetic transition frequencies, which lie at the heart of the TGD based theory of brain consciousness. Classical fields affect various macroscopic quantum phases and one cannot exclude the possibility that gene level is involved with cognition and sensory experiencing even in the time scales shorter than the long time scales associated with the neural transmitter action.

Neural level could control genetic level (this is known to occur chemically) via non-chemical control mechanisms. Nuclear matrix, cytoskeleton and the collagen network associated with connective tissues are liquid crystals and ideal tools for transforming electrical signals to mechanical (say conformational wave s) and vice versa. The massless extremals associated with the micro-tubules connecting neuronal cell membrane with the nuclear region could make possible the information transfer from neural level to gene level: in fact, the non-dispersive vacuum currents associated with the massless extremals are optimal for communication purposes. This suggests that nerve pulse patterns could be transferred into genome along cellular matrix.

The idea that information from the genetic level would be transferred to neural level and that genetic level could even control neural level chemically is consistent with experimental facts. One could even consider the possibility that genome plays the role of neuronal brain. One can indeed play with idea that possible "this is true" experiences associated with the active genes could be expressed as our emotions. As noticed, the time scale of chemical gene expression is .05 seconds and is much slower than the millisecond time scale of nerve pulse so that direct translation of genes to nerve pulse patterns is not plausible. This is however consistent with the realization of "this is true" experiences as emotions which are characterized by slow time scales. Thus the statements "If A then B" expressed by genes B and the control structures A associated with them could give rise to neural activity translating these experiences to sensory experiences about the internal state of brain accompanied by and often identified with emotions, which in turn could be expressed as internal speech. Music metaphor does not however require any precise coding to nerve pulse patterns. The idea that this communication could occur also non-chemically is much more speculative and not actually supported by the most stringent form of the master slave hierarchy. It is also difficult to imagine how the coding of DNA sequence to, say nerve pulse patterns, could be achieved. In principle, memetic code could be realized as sequences of 21 DNA triplets. The idea about direct one-to-one mapping of DNA level to memetic level seems however implausible.

The large amount of silent DNA suggest that some non-chemical gene expression mechanisms might be at work. This does not necessitate communication between genomes although also this is possible in principle. The extreme would be that neuronal genomes could form neuronal democracy communicating to each other their "If A then B" opinions with conclusion B depending on what the exon-intron decomposition of gene B is for particular gene. This would make possible neuronal voting. Assume that the sub-selves of self are neuron groups with identical synchronized inputs A. This means that same genes are activated and output depends only on the intron-exon

decomposition of the individual gene and that sub-self experiences entire spectrum of opinions about what A implies. Self experiences the neuron group sub-self as the average over the opinions "If A then ..." of individual genome subsub-selves. One can also consider the possibility of Boolean "machines". Gene B receives the premises A of "If A then B" as nerve pulse pattern acting to the control sequence associated with B and generate as output the nerve pulse pattern representing statement B when premises are satisfied.

1.4 Model For The Genetic Code

The basic numbers of genetic code are probably not accidental. This led for ten years ago to an attempt to construct a model for abstraction process reproducing the basic numbers of the genetic code. The simplest model for an abstraction process is based to a repeated formation of statements about statements starting from two basic statements. If one drops at each step of construction the statement corresponding to empty set in the set theoretic realization of Boolean algebra, one obtains a hierarchy allowing to understand the basic numbers of genetic code.

What one obtains is so called Combinatorial Hierarchy [**?**]onsisting of the Mersenne numbers $2, M(1) = 3, 7.127, 2^{127} - 1, ..$ constructed using the rule $M(n+1) = M_{M(n)} = 2^{M(n)} - 1$. The explicitly listed ones are known to be primes. Combinatorial Hierarchy emerges from a model of abstraction process as subsequent transitions from level to meta level by forming Boolean statements about Boolean statements of level n and dropping one statement away. Combinatorial Hierarchy results also by constructing the sets of all subsets with empty set excluded starting from two element set. The set of statements at level n can be given a structure of Finite Field $G(M(n), 1)$ if $M(n)$ is prime. The multiplicative groups $Z_{M(n)-1}$ form a nested hierarchy and the coset spaces $Z_{k_n} \equiv Z_{M(n+1)-1}/Z_{M(n)-1}$ are cyclic groups. Combinatorial Hierarchy based model of Genetic Code explains the number of DNA:s and amino-acids and the representation of words of the GC as triplets of 4 different codewords. Aminoacids correspond to $k_{n=3} = 21$ axioms of a formal system defined by $n = 3$ level of Combinatorial Hierarchy having a unique imbedding as the group $Z_{k_n} \subset Z_{M(n)-1} = Z_{126}$ and DNA:s correspond to the set $X_{N(DNA)} \subset Z_{M(n)-1}$ of $N(DNA) = (M(n)+1)/2 = 64$ statements consistent with a given atomic statement at level n regarded as special cases of general theorems. GC corresponds to the mapping $x \to x^{k_{n-1}} = x^6$ in $Z_{M(n)-1}$ mapping DNA type statements to amino-acid type statements. The numbers of DNA:s coding single amino-acid are reproduced in a symmetry breaking mechanism involving the finite groups $Z_{p_{n-1}}$ and Z_{k_n} and symmetry breaking is in a well defined sense minimal. The infinite hierarchy of possible genetic codes suggests the possibility of an infinite hierarchy of increasingly complicated life-forms.

The physical model of the genetic code leads to a beautiful interpretation of the genetic code as mapping the fundamental 64 truths to 20 basic conscious experiences, perhaps the emotion about truthness. The fact that hormones correlate with emotions suggests that this map assigns to a logical statement an emotion and that even our emotions could relate to DNA level. Note however that TGD predicts entire hierarchy of selves. Aminoacid P corresponds to the emotionally experienced truth $G_1(P)$ or $G_2(P)$.. or $G_n(P)$ is true, where G_i code for protein P. 3 stopping sign codewords cannot be experienced emotionally as truths not non-truths (holy trinity!).

1.5 The Relationship Between Genetic And Memetic Codes

TGD leads to a model of Boolean mind in terms of the temporal sequences formed by cognitive neutrino pairs with vanishing total energy. As noticed, the model for abstraction process predicts entire hierarchy of genetic codes [**?**] This leads to the idea that our cognition might correspond to the level next to the genetic code. The hypothesis that memetic code corresponds to the next level of Combinatorial Hierarchy characterized by Mersenne prime $M_{127} = 2^{127} - 1$, when combined with p-adic length scale hypothesis, leads to a prediction of about .1 seconds for the duration of the "wake-up" period of sub-self corresponding to the codeword of the memetic code.

The memetic codeword consists of 126 bits and could be represented by two possible spin directions of fermion corresponding to two values of Boolean statement. This implies that one millisecond should be the duration of single bit: this time scale is indeed fundamental for nerve pulse activity. What the fermions in question are? Ths question must be left open. The original proposal based on cognitive neutrino pairs at opposite throats of wormhole contact discussed

in [?]ooks highly unrealistic from the point of view of standard model physics. The possibility of scaled variant of weak interaction physics with intermediate boson length scale of order cell size or cell membrane thickness however allows to consider this kind of option. Dark protons is another assignable to the lipid layers of cell membrane is second option [?]

This picture suggests the following general framework. Memes and genes correspond to two levels in the hierarchy of conscious intelligence and genetic programs could be perhaps seen as subprograms called by the higher level memetic programs. One could even see higher level life as symbiosis of memes and genes.

This raises several questions. Does genetic code determine the evolution of the organism or is the development of organism some kind of "social process" in which the genetic level interacts with the memetic level? Do genes code the space-time sheets representing the hardware of the memetic level or do higher level organisms represent symbiosis of two almost independent life forms? Does memetic level control genetic level or vice versa or is the interaction between these levels bi-directional? Despite that these questions which remain open, it seems that the three approaches to understand cognition based on the Combinatorial Hierarchy model of abstraction process, to fermionic Boolean algebra as a model of logical mind and to genetic and menetic programs as a model for conscious intelligence seem to combine to form single "holy trinity" of cognition.

1.6 Mersenne Hypothesis

The hierarchy of dark matter levels is labeled by the values of Planck constant having quantized but arbitrarily large values TGD inspired quantum biology and number theoretical considerations suggest preferred values for $r = \hbar/\hbar_0$. For the most general option the values of $\hbar$ are products and ratios of two integers n_a and n_b. Ruler and compass integers defined by the products of distinct Fermat primes and power of two are number theoretically favored values for these integers because the phases $exp(i2\pi/n_i)$, $i \in \{a, b\}$, in this case are number theoretically very simple and should have emerged first in the number theoretical evolution via algebraic extensions of p-adics and of rationals. p-Adic length scale hypothesis favors powers of two as values of r.

The hypothesis that Mersenne primes $M_k = 2^k - 1$, $k \in \{89, 107, 127\}$, and Gaussian Mersennes $M_{G,k} = (1+i)k - 1$, $k \in \{113, 151, 157, 163, 167, 239, 241..\}$ (the number theoretical miracle is that all the four scaled up electronic Compton lengths with $k \in \{151, 157, 163, 167\}$ are in the biologically highly interesting range 10 nm-2.5 μm) define scaled up copies of electro-weak and QCD type physics with ordinary value of $\hbar$ and that these physics are induced by dark variants of corresponding lower level physics leads to a prediction for the preferred values of $r = 2^{k_d}$, $k_d = k_i - k_j$, and the resulting picture finds support from the ensuing models for biological evolution and for EEG [?] This hypothesis - to be referred to as Mersenne hypothesis - replaces the earlier rather ad hoc proposal $r = \hbar/\hbar_0 = 2^{11k}$ for the preferred values of Planck constant. The background necessary for understanding what is involved is described in [?]

1.7 Fractal Hierarchy Of Magnetic Flux Sheets And The Hierarchy Of Genomes

The notion of magnetic body is central in the TGD inspired theory of living matter. Every system possesses magnetic body and there are strong reasons to believe that the magnetic body associated with human body is of order Earth size and that there could be an entire hierarchy of these bodies with even much larger sizes. Therefore the question arises what one can assume about these magnetic bodies. The quantization of magnetic flux suggests an answer to this question.

1. The quantization condition for magnetic flux reads in the most general form as $\oint (p - eA) \cdot dl = n\hbar$. If supra currents flowing at the boundaries of the flux tube are absent one obtains $e \int B \cdot dS = n\hbar$, which requires that the scaling of the Planck constant scales up the flux tube thickness by r^2 and scaling of B by $1/r$. If one assumes that the radii of flux tubes do not depend on the value of r, magnetic flux is compensated by the contribution of the supra current flowing around the flux tube: $\oint (p - eA) \cdot dl = 0$. The supra currents would be present inside living organism but in the faraway region where flux quanta from organism fuse together, the quantization conditions $e \int B \cdot dS = n\hbar$ would be satisfied.

2. From the point of view of EEG especially interesting are the flux sheets which have thickness $L_e(151) = 10$ nm (the thickness of cell membrane) carrying magnetic field having strength of endogenous magnetic field. In absence of supra currents these flux sheets have very large total transversal length proportional to r^2. The condition that the values of cycloctron energies are above thermal energy implies that the value of r is of order 2^{k_d}, $k_d = 44$. Strongly folded flux sheets of this thickness might be associated with living matter and connect their DNAs to single coherent structure. One can of course assume the presence of supra currents but outside the organism the flux sheet should fuse to form very long flux sheets.

3. Suppose that the magnetic flux flows in head to tail direction so that the magnetic flux arrives to the human body through a layer of cortical neurons. Assume that the flux sheets traverse through the uppermost layer of neurons and also lower layers and that DNA of each neuronal nuclei define a transversal sections organized along flux sheet like text lines of a book page. The total length of DNA in single human cell is about one meter. It seems that single organism cannot provide the needed total length of DNA if DNA dominates the contribution. This if of course not at all necessarily since supra currents are possible and outside the organism the flux sheets can fuse together. This implies however correlations between genomes of different cells and even different organisms.

These observations inspire the notion of super- and hyper genes. As a matter fact, entire hierarchy of genomes is predicted. Super genes consist of genes in different cell nuclei arranged to threads along magnetic flux sheets like text lines on the page of book whereas hyper genes traverse through genomes of different organisms. Super and hyper genes provide an enormous representative capacity and together with the dark matter hierarchy allows to resolve the paradox created by the observation that human genome does not differ appreciably in size from that of wheat.

The appendix of the book gives a summary about basic concepts of TGD with illustrations. There are concept maps about topics related to the contents of the chapter prepared using CMAP realized as html files. Links to all CMAP files can be found at `http://tgdtheory.fi/cmaphtml.html` [L4]. Pdf representation of same files serving as a kind of glossary can be found at `http://tgdtheory.fi/tgdglossary.pdf` [L5]. The topics relevant to this chapter are given by the following list.

- Magnetic body [L8]
- Basic Mechanisms associated with magnetic body [L3]
- Genes and memes [L7]
- Dark proton strings and genetic code [L6]
- Nuclear string model [L9]
- Origin of genetic code [L10]
- Pollack's observations [L11]

2 Background

The foundations of genetics were discovered by George Mendel in 1866, but remained generally unknown until 1900. During the first half of nineteenth century it was gradually realized that genes play major roles in the functioning and evolution of organisms. The discovery of DNA revealed the principles of heredity and how genes store hereditary information and transmit it from generation to next. Hereditary information is contained within the nucleotide sequence of DNA.

Organization, expression, and evolution of the hereditary information are the main aspects of genetics. Hereditary information is organized into chromosomes consisting of DNA sequences. It is expressed via transcription to mRNA followed by a translation to protein. The evolution of the hereditary information involves basically sexual breeding in one chromosome from the chromosome pairs of both parents combine to form chromosome pair. Also recombination of the members of

the chromosome pairs is possible during meiosis. Also other mechanisms, such as fusion or fission of chromosomes and modification of DNA sequences, are possible. There are excellent books about topics [I3] but for the convenience of the reader the basics of genetics are very briefly summarized in the following.

2.1 DNA And RNA

DNA add RNA provide a manner to store and organize genetic information [I3].

1. Genetic information is stored in nucleic acids, which are long sequences of nucleotide serving as letters of genetic code: three nucleosides form single world of code. There are four different nucleotides so that the number of different words is 64.

2. Nucleotide consists of three basic units joined by covalent bonds: nucleotide= nucleoside+sugar+$5'$-phosphate. The units are sugar, which is deoxiribose in case of DNA and ribose in case of RNA, phosphate and nucleoside (nucleic acid). Nucleosides are the information carrying part of DNA and RNA.

3. DNA and RNA sequences contain 4 different nucleosides. In case of DNA they correspond to C(ytocine), T(ymine), A(denine) and T(ymine). In case of RNA T is replaced by U(racil). U, T and C are purines containing one carbon ring and A and G are pyrimidines containing two carbon rings.

4. DNA molecules/nucleic acids/polynucleotides are formed as very long sequences of nucleotides bound together by phospho-diester bonds.

DNA double helix consists of two DNA strands, which are conjugates of each other, conjugation being defined as $A \leftrightarrow T$, $C \leftrightarrow G$. The helices are bound by hydrogen bonds between A and T and C and G respectively.

Sequences of DNA triplets form genes, which represent basic units of hereditary information revealed as traits of the organism. Each gene involves also additional DNA sequences serving as control structures in the transcription of gene to mRNA. In prokaryotes there is only single chromosome in the form of a circular double strand. In eukaryotes the chromosomes are located in nucleus and appear in homologous pairs. Eukaryotic chromosome is a complicated helical structure resembling beads in thread formed by DNA. DNA is wound around nucleosomes with diameter $d \simeq 10$ nanometers. Nucleosomes consist of octamer formed from 4 different histones. Chromosome structure will be considered in more detail later.

RNA appears both inside nucleus and cell. There are several types of RNA.

1. Messenger RNA (mRNA) is the outcome of transcription of DNA inside nucleus and is translated to proteins outside the nucleus.

2. Transfer RNA (tRNA) is involved in the translation of mRNA to protein: tRNA molecules bind specific amino-acids and glue them to specific mRNA triplets in a manner dictated by genetic code. rRNA appears as a building block protein of ribosomes playing the role of reading head in the translation of mRNA to proteins.

3. In case of eukaryotes transcription involves intermediate state in which DNA is transcribed to hnRNA which contains also the transcriptions of introns ("junk DNA"), which are split in so called splicing process cutting away intron RNA to form RNA-protein complexes which remain inside nucleus.

2.2 Proteins

Proteins are in a vital role in organisms. The diversity and complexity of life is largely due to the diversity and complexity of proteins. Some proteins act as transcription factors controlling genetic expression. Some proteins are used by cells in chemical communication between cells: hormones serve as signalling proteins; various receptor proteins serve as receptors of chemical signals and hormone-receptor complexes serve as transcription factors. Neural transmitters appear in the

synaptic communication between neurons. Some proteins act as enzymes catalyzing biochemical reactions. Other proteins serve as structural building blocks, either by themselves or in association with nucleic acids (nucleoproteins), polysaccharides (glycoproteins) or lipids (lipoproteins). Some proteins, such as myoglobins and hemoglobins are associated with metal-containing organic molecules.

Proteins consists of polypeptides, which are polymers of 20 different amino-acids. Genetic code assigns unique polypeptide to a given gene. With single exception amino-acids share the same basic structure. Hydrogen atom H, carboxyl group COOH and amino group NH_2 and radical R linked to carbon atom. R determines exclusively the chemical properties of protein. 8 amino-acids are nonpolar (hydrophobic) and 12 of them are polar (hydrophilic). Of the twelve polar amino-acids 7 are neutral, 3 are basic (tending to become positively charged) and 2 and acidic (tending to become negatively charged) under physiological conditions. Carboxyl and amino groups tend to become ionized at physiological pH; -COOH group tends to lose its proton and NH_2 group tends to gain a proton.

In polypeptides, which are products of gene translation, amino-acids are linked to each other by peptide bonds formed when NH_2 group of one amino-acid and $COOH$ group of next amino-acid are linked (H_2O molecule is snipped away in this process). Polypeptide chains spontaneously adopt so called secondary structure determined by the nature of the R groups along the backbone. Backbone forms alpha helix, a coil containg 3.6 amino-acid redidues per turn. Another secondary structure is the beta pleated sheet configuration consisting of rows of polypeptide chains hydrogen bonded with each other. Polypeptide can also adopt the form of a random coil. Proline, because of its unique structure, causes a kink in the polypeptide backbone. Polypeptides have also tertiary structure. How the tertiary structure is determined by the chemistry of amino-acids is poorly understood. One of the big problems of biology is to understand who protein is able to fold to such a unique configuration. TGD suggests that tertiary structure might not be determined solely by the standard chemistry and that many-sheeted nature of protein might be crucial in determining the final result of the folding. There is also quaternary structure associated with proteins formed by polypeptide sequences. The formation of higher level structures, such as micro-tubules, micro-filaments, cell membranes and collagen fibers involves self-organization and living matter seems to behave as a liquid crystal whose basic properties depend only on very general properties of protein.

2.3 Replication, Transcription, Translation

Information processing in living matter involves three basic processes: replication, transcription and translation. Replication of DNA means replication of DNA double helices and is essentially copying of genetic information. Replication involves unwinding of the parental strands of DNA double helix. They serve as templates on which the growing complementary daughter strands are synthesized. The direction of the synthesis is opposite for the two strands and only the second (leading) strand can be synthetized continuosly whereas the synthesis of the second strand occurs discontinuously and results in disjoint pieces of DNA containing approximately 1000 nucleotide pairs (Okazaki fragments of length about 34 nanometers), which later combine to form connected DNA strand.

DNA can be transcribed to mRNA molecules (messenger RNA) translated to proteins; to tRNA (transfer RNA), which is the RNA molecule affecting the coding of RNA triplets to amino-acids and to rRNA, which is the building block of the machinery affecting the translation. In case of prokaryotes the transcription of DNA to mRNA occurs directly. The rules for the transcription are

$$\begin{aligned} A \to U \ , \quad T \to A \ , \\ C \to G \ , \quad G \to C \ . \end{aligned}$$

In case of eukaryotes the transcription involves two steps since eukaryote genes in general decompose into exons translated to protein plus introns.

First the entire gene is transcribed to hnRNA sequence. After this so called splicing occurs and gives rise to mRNA, which corresponds to the DNA sequence formed by the exons. In the splicing process intron sequences are split off and wind around specific proteins which do not leave the nucleus. There are different pathways for slicing meaning that the decomposition to exons and

introns is not unique. Dynamical exon-intron decomposition is essential for the working of immune system.

Transcription is a complicated process involving the action of several enzymes. RNA polymerase I is involved in the transciption of large rRNA molecules, RNA polymerase II with the transcription of hnRNA, RNA polymerase III with transcription of small 5S-rRNA molecules and tRNA molecules. Usually so called heavy strand is transcribed. Light strand can be transcribed to some tRNA molecules at least. Gene is preceded by AUG triplet. In eukaryote cells RNA polymerase II copies sequences containing 6000-8000, sometimes even 20.000 nucleotides. The average length of mRNA sequence is 1500 nucleotides and the amino-acid corresponds to a sequence of average length of 1200 nucleotides.

RNA II polymerase binds to the promotor region preceding the gene. Promotor region contains at least two binding sites, so called TATA block and CCAT sequence regognized by RNA polymerase. Between promotor site and gene are operator site in which repressor enzymes bind and make translation impossible. TAC sequence denotes the beginning of that part of gene which is translated to protein (apart from introns). At the end of the gene there is rather long $A \cdots AAA$ control sequence preceided by TGA sequence signifying the end of the part of the gene to be translated. Introns which are not translated begin with AC and end with CA.

The translation of mRNA to polypeptide occurs outside the nucleus. Translation involves tRNA molecules, which are about 80 nucleotides long. Each tRNA contains a specific triplet which is anticodeword for the corresponding codeword in mRNA and binds only to this codeword in translation process. Each tRNA molecule binds with a specific amino-acid molecule and each amino-acid has at least one tRNA binding to it. The allowed bindings of tRNA and amino-acid molecules define genetic code. In translation tRNA carrying amino-acid attaches to an mRNA codeword to its own anticodeword and the amino-acid forms a peptide bond with the polypeptide sequence already translated at rRNA.

Genetic code assigns to 64 RNA triplets 20 amino-acids so that there is a considerable degeneracy involved. The largest number of DNA codewords mapped to same amino-acid is six. Three codewords are interpreted as stoppping sign for the translation. Genetic code is universal for the nuclear DNA of all eukaryotes and prokaryotes. The mitochondrial genetic codes of various eukaryotes however differ slightly from the universal genetic code. For instance, 4 DNA triplets can correspond to stopping sign.

Replication, transcription and translation are not the only information transfer processes occurring in living matter.

1. Reverse transcription $RNA \to DNA$ is known to occur in some cases and is also involved with the homing phenomenon of introns. Reverse transcription might have led from a system of RNA and proteins to system involving DNA sequences and primitive form of genetic code. The simplest starting system of this kind would consist of DNA coding RNA coding a protein which catalyzes both transcription and reverse transcription. This kind of system might have gradually evolved to a more complex DNA sequences.

2. RNA replication can occur in cells infected by viruses. What happens is that viral RNA strand which can be either single or double stranded, is replicated to its complement which in turn serves as a template for the synthesis of progeny RNA molecules.

3. Direct translation of DNA to protein without transcription has been observed only in vitro. This process probably never occurs in living cells.

2.4 Introns, Pseudogenes, Repetitive DNA, Silent DNA

The genes in nuclei of the eukaryote cells contain introns, sequences consisting of 10-1000 nuclotides interspersed with the exon parts of DNA which is translated to a protein coded by gene [I2] . Molecular Darwinist could compare introns with the commercials appearing between TV program or simply as selfish DNA. One could see them also unused parts of a computer program separated from the program code by comment signs in front of each line corresponding now to DNA nucleotide. The latter metaphor is consistent with the observation that intron can begin even in the middle of DNA triplet and that the transcription to mRNA is not unique so that same gene can

give rise to several proteins. The content of intron sequences seem to be unrelated to the exon sequences: as if two separate interspersed computer codes would be in question.

Only one prokaryote cell, photosynthetic cyanobacterium Fischerella, is known to contain introns [I2]. Usually also the genes of cell organelles (such as mitochondria of human cell) contain only very few introns. Fungi are however an exception in this respect [I3]. The higher the evolutionary level of the eukaryote cell, the higher the fraction of introns in the genome is. For humane genome the fraction of exons is about one per cent. During transcription both introns and exons are transcribed to hnRNA, intron sections are snipped away in a process called splicing and the resulting mRNA for the protein coded by exons is transferred from the nucleus and translated to a protein coded by the gene. It is possible to snip off the introns from genome but the mRNA coded by these genes is not transferred from nucleus, which suggests that introns have some role in genetic program. The addition of introns does not seem to have any dramatic effects on the genetic program.

Introns are a headache of molecular Darwinism. The nickname "junk DNA" tells the basic attitude towards introns. Introns represent selfish DNA living as parasites of the genome. There are two opposite schools concerning how introns have appeared.

1. The first school claims that introns came early. Somewhat surprisinly, this school sees bacteria as results of a long evolution which has gradually snipped off the introns from a primitive cell in order to achieve maximal rate of DNA transcription. One can of course wonder why the same thing has not happened to the cell nuclei also.

2. Second school tells that introns came late: this view conforms with the observations about the fraction of introns in genome. Introns seem to start from preferred sites and exons seem often to correspond to a modular decomposition of the protein they code. On basis of this it has been also proposed that introns separate modular parts of proteins from each other. The facts that introns can appear in the middle of protein module and even split single DNA triplet are not however consistent with this interpretation.

One can criticize the identification of introns as junk DNA.

1. It is difficult to see how human genome containing so high per cent of junk DNA could work with such a fantastic precision while viruses, second form of junk DNA, are often lethal. There are several pathways for slicing. Exon-intron transformation has been found to occur: exon and intron parts of gene simply change their roles [I2, I5] ! This suggest that exon-intron property is additional dynamical degree of freedom in genone and might have deep meaning. Exon↔exon transformation is indeed crucial for the working of immune system.

2. mRNA produced by intronless gene does not get out of nucleus. It seems that the presence of introns somehow initiates a module of genetic program taking care that protein mRNA gets out of the nucleus. Introns seem thus to be necessary for the functioning of the cell and could be in some sense regarded as an output of gene interpreted as a genetic subprogram. Note however that intron mRNA which winds around spherical proteins in the process of splicing, have not been reported to serve as transcription factors.

3. The positions of the intron sequences in similar genes are not same for various species. There are wandering introns which can move even from cell to another one. There is a phenomenon called homing [I2, I6]: the RNA coded by intron inserts itself into DNA sequence and builds by inverse transcription its complement in complementary DNA strand. In retrohoming intron transforms to RNA installed to DNA sequence by reverse splicing [I4] . This suggests that introns might provide a new mechanism for the evolution of the genome and provide a mechanism for modifying the program code of genetic programs. It is also known that there are long range correlations (in scale of one micro-meter) in genes containing introns [I7]. This suggests that introns are essential element in the organization of DNA to larger structures.

All these properties of introns suggest that their role in genetic program is badly misunderstood in the framework provided by molecular Darwinism and the basic dogmas of genetics.

Besides introns there are pseudogenes of various types, which by definition code no proteins. For instance, eukaryote genes from which introns have been snipped off, behave as pseudogenes.

Pseudogenes can also contain "programming errors": for instance, the DNA triplet signifying the beginning of gene has changed. Genetic program metaphor suggests interpretation of pseudogenes as unused program modules. The idea about two interspersed program codes could explain the program errors as only apparent program errors. Of course, every experienced computer programmer would suggests the possibility of also genuine program errors! Also the interpretation as control structure affecting transcription via long range interactions rather than via chemical contact interactions might make sense. It is indeed known that so called enhancers and silencers act as transcription factors in this manner.

Genetic code contains large amounts of repetitive DNA.

1. Five per cent of genome of the eukaryotes consists of highly repetitive DNA consisting of 5-300 nucleotides (even 10^6 copies are possible). In particular satellite DNA, containing less than 10 nucleotides belongs to this class. This DNA are active during mitosis and meiosis [I3].

2. 30 per cent of DNA is moderately repetitive. The first class corresponds to rRNA, 5SRNA, tRNA and histogenes (10-100 copies). These genes are concentrated in certain chromosomes. In case of genes coding rRNA, tRNA the repetition of genes is understandable since translation making possible large number of amino-acid copies does not occur. The fact is however that also genes coding proteins appear as very many copies and there is no obvious explanation for this. So called SINE segments have length not longer than 10^3 np and are interspersed through the entire genome as $10^4 - 10^5$ copies. LINE-segements consist about 3×10^3 np: there are about 10^4 copies are interspersed through the entire genome. Part of these sequences are transposons (see below).

3. 65 per cent of DNA are present in only few (1-15) copies. Both exons and introns belong to this group of DNA and exons form only one percent of human genome.

4. The control regions between genes are rather long and seem to contain DNA with no obvious function. Also second strand of DNA can be regarded as silent DNA since its presence is not absolutely necessary for the storage of genetic information. The question is whether this silent DNA has some hitherto unidentified function.

The genome of both prokaryotes and eukaryotes contains transposons, which are movable DNA sequences able to insert themselves to DNA with the help of insertion sequences. Insertion sequences are short (less than 2000 nucleotides) and do not code proteins. Insertion sequences can carry also promotor and repressor sequences with them. Transposons could be important for evolution.

2.5 Is Central Dogma An Absolute Truth?

The Central Dogma of molecular biology states that each gene corresponds to a unique polypeptid. There are several observations challenging Central Dogma.

1. It is known that many alternate pathways of transcript splicing are possible and give rise to different protein outcomes called isoforms. This would suggest that transformation of some introns to exons and vice versa occurs routinely in gene expression. Using computer program analogy, this transformation would mean that the program part represented by introns becomes active and the part represented by exons becomes passive.

2. The phenomenon of superimposed genes [I3]. There are genes nested inside genes and translation can start also in the middle of gene producing shorter protein than the gene usually. These phenomena were first observed for bacteriophage $\phi X174$, whose genome is known in its entirety. It is known that gene is transcribed as a whole and that different proteins result from frame shift. Gene can also overlap the DNA sequences formed by two subsequent genes as first observed in bacteriophage $G4$. These observations suggest that the standard notion of gene fails somehow.

3. It is known that also the "nonsense" strand of DNA can serve as template for transcription [I3].

Table 1: The amount of total genome measured as the number of DNA triplets.

Organism	Human	Mus musculus	Amoeba	Marbled lungfish
$N(DNA)/10^9$	3	3	670	139
Organism	Salamander	Onion	Trumpet lily	
$N(DNA)/10^9$	81	18	90	

2.6 Is Life Nothing But Biochemistry?

It is not at all obvious whether the hypothesis "life is nothing but biochemistry" holds true.

1. It is not known whether protein folding is coded into the chemistry of DNA. The problem is mathematically unattractable due to the occurrence of combinatorial explosion. It seems more probable that folding might be self-organization type phenomenon and thus affected by the conditions of environment: protein development can be regarded as hopping in spin glass type energy landscape leading to some deep valley of free energy valley. TGD suggest that folding is the quantum analog of this kind of process. In particular, p-adic length scale hierarchy and many-sheeted space-time concept (see **Fig.** `http://tgdtheory.fi/appfigures/manysheeted.jpg` or **Fig.** 9 in the appendix of this book) suggest that one cannot understand protein folding in terms of DNA chemistry alone.

2. DNA is essentially one-dimensional structure. This suggests that gene codes only one-dimensional skeleton of its expression domain and that self-organization by quantum jumps could take care of the rest. Indeed, the work of Mae-Wan Ho [I8] shows that living organisms are liquid crystals which can be regarded as one-dimensional crystals and two-dimensional liquids, whose properties can be characterized by some global parameters. Perhaps genes code the properties of various layers of the liquid crystal. One of the basic characteristics of liquid crystals is self-assembly and de-assembly. Depending on pH, ionic concentrations, temperature, electric fields, ... liquid crystals organize to micelle like structures (cell membranes, collagen fibers, ...) and effectively one-dimensional layered structures [D1].

3. One can wonder how morphology is coded in DNA and how it is decoded from DNA. It is not at all obvious that DNA chemistry, which is purely local, is enough to code morphology.

4. So called enhancers and silencers are transcription factors, which encourage or disencourage gene expression in eukaryotes. The position of these proteins or orientation in DNA does not seem to be important [I3]. For instance, they can bind to introns and the distance of the binding site from gene promotor regions can be thousands of nucleotide pairs. This would suggest that the mechanisms of enhancing and silencing are not purely chemical if chemical at all. This would suggest the generalization of the notions of gene expression and transcription factor. Chemical expression takes place very slowly. Non-chemical expression modes yielding non-chemical transcription factors could make possible very fast running of genetic programs and there could be even connection between many-sheeted genome and nerve pulse activity.

5. The naive expectation is that the size of the genome should correlate with the evolutionary stage of the species. Eukaryotes indeed have genome which is typically 10^3 times longer than prokaroyote genome. The **Table 1** however shows that the total length of genome does not correlate with the complexity of the organism faithfully. The genome of plants is typically 10-100 times longer than human genome. The genome of amoeba is by two orders of magnitude longer than that of human! The genomes of monkeys and men are almost identical. This suggests that there might be some unindentified degree of freedom associated with DNA which explains these differences.

3 Many-Sheeted DNA

The notion of many-sheeted DNA suggest a profoundly new manner to understand how the morphology of the organism is coded to and decoded from DNA. p-Adic length scale hypothesis leads to precise quantitative predictions for the number of levels of genetic program as function of a suitably defined size of the organ. The proposed model for introns leads to the interpretation of gene as a representation for Boolean algebra and to the proposal that genes realize not necessarily conscious-to-us Boolean mind at the basic level. What is especially nice is that connection with the realization of Boolean mind in terms of cognitive neutrino pairs is consistent with this picture. Many -sheeted DNA suggests also new forms of gene expression and of control of gene expression. For instance, nerve pulse patterns could affect also genetic program of postsynaptic cell via the classical em and Z^0 field patterns associated with them and genes could affect cell membrane via conformational waves propagating along micro-tubules connecting nucleus to cell membrane.

3.1 Many-Sheeted DNA As Hierarchy Of Genetic Programs

Many-sheeted DNA allows to realize genetic subprogram hierarchy in an elegant manner. Many-sheeted DNA and proteins are like a hierarchy of ordinary DNA and proteins effectively living in different space-times corresponding to body parts. One can consider the possibility that subprograms correspond to p-adic space-time sheets and subprogram hierarchy corresponds to the hierarchy of p-adic space-time sheets. The gene program in a given length scale would selectively activate programs in shorter length scale, etc.. DNA sequences with the same chemical structure correspond to different genetic programs since the many-sheeted structure of DNA affects its functioning. Analogous conclusion is true about proteins.

One can assign to gene a unique p-adic prime as the prime characterizing the largest p-adic sheet at which gene has # contacts. The number of levels in subprogram hierarchy could be deduced from the size of the organism. Gene can have # contacts to several space-time sheets characterized by p-adic primes $p \simeq 2^k$, k power of prime. Denote by k_G the largest value of k associated with gene. k_G characterizes the position of gene in subprogram hierarchy. Gene can have # contacts with space-time sheets $k < k_G$ also. Gene can be characterized by the p-adic k_G labelling the largest space-time sheet to which it has # contacts. "Comment sign" marking each nucleotide of intron could correspond to a direction of classical field at some space-time sheet characterized by p-adic prime $p \simeq 2^k$, $k = k_I$. The only sensible assumption seems to be $k_I = k_G$.

The other # contacts of gene must be assumed to be on space-time sheets with $k < k_G = k_I$. This implies that given program can call only programs which are in the lower level of the hierarchy. This would suggest that programs belonging at the lower level of hierarchy cannot call program at higher level. Does this imply that growth process in which larger and larger space-time sheets are activated can only occur by self-organization? This would mean that DNA space-time sheets with increasing value of k_G expand in phase transition like manner and fuse to form space-time sheets corresponding to various body parts. On the other hand, it is not at all obvious that growth process could not start from higher level and lead to gradual differentiation at lower levels. In fact, embryogenesis seems to occur in this manner [I3].

Also proteins can be classified by the the number k_P characterizing the largest space-time sheet to which protein has # contacts. Proteins must mediate program calls to gene modules G_1 with various values of $k_{G_1} < k_G$. This suggests that protein activating gene characterized by k_{G_1} must have same $k_P = k_{G_1}$. This would automatically guarantee that chemically identical proteins activate only the genes belonging to the level of the fractal hierarchy they represent.

The notion of many-sheeted DNA has immediate applications.

1. Many-sheeted DNA provides a possible explanation for why DNA triplets act as codewords of the genetic code. If members of each DNA triplet are glued to space-time sheet containing only # contacts from the nucleotides of the triplet, codewords have a clear geometrical meaning.

2. The notions of many-sheeted DNA and many-sheeted protein suggests also an explanation for how enhancers and silencers are able to regulate gene expression. Interaction with classical em or Z^0 fields via wormhole contacts provides a non-chemical interaction mechanism. Second mechanism is based on Josephson currents running along join along boundaries contacts.

Since interaction with much larger length scale is involved, these interaction mechanisms are not too sensitive to the position of the transcription factor and the distance of the binding site from gene promotor regions can be thousands of nucleotide pairs. This mechanism explains also the observe issue specificity of some transcription factors. Proteins with same chemical structure can be quite different transcription factors if they have contacts to different space-time sheets.

3.2 Possible Answers To The Basic Questions

Many-sheeted DNA suggests stupifyingly simple coding of body's morphology. The genes would be obtained by simply contracting the many-sheeted space-time representing expression domains of genes to one-dimensional structure. Decoding of the morphology means the growth of this structures to their original size. Of course. this hypothesis is oversimplified but its extreme simplicity makes it worth of testing.

3.2.1 How the structure of expression domain of the gene is coded in the structure of gene?

The p-adic length scale of the gene correlates trivially with the p-adic length scale of the protein coded by it. Already protein folding implies that the correlation with the size of the structure coded by DNA is not so straightforward. Furthermore, proteins are not mere building blocks but can have quite abstract functions like regulating gene expression of genes.

Consider now various aspects of the idea that expression the domain of gene is coded into the structure of gene and this that correspondence could be also realized at functional level.

1. The first thing that comes into mind is that the p-adic length scale of the gene correlates with the p-adic prime of the space-time sheet which corresponds to the expression domain of the gene during early phases of the embryogenesis. Gene clusters, say Hox cluster, would represent kind of a miniature of the body and every gene of Hox cluster would give rise to a space-time sheet which would be a scaled down model of the expression domain of the gene. Thus the expression domains of various genes in the genome could correspond to the extended space-time sheets at the level of the genome and the topology of these genome level expression domains, in particular, their ordering, would be consistent with that for the actual expression domains. Expression domain corresponds most naturally to a join along boundaries condensate generated by the formation of the join along boundaries bonds between the extended space-time sheets associated with the genes. This means that the p-adic prime of the expression domain can be much smaller than one could conclude it to be on basis of its size.

2. One could test the hypothesis that the total length of the region occupied by gene and by the DNA controlling its activity in the genome could correlate with the size of its expression domain at the stage of the development when the gene is expressed. Note that many genes affecting morphogenesis are expressed in a very early stage: many of them in the embryonic stage when no cell formation has yet occurred. This stage corresponds to the p-adic length scale of a fertilized cell about 10^{-4} meters. Of course, the correlation between the content of the gene program and the size of its expression domain, is not necessary and might be even un-desirable.

3. Fractality suggests that the communication by expression factor proteins at the level of genome might mimic the hormonal communication occurring at the level of the entire organism. This could mean that the hormonal communication between the expression domains of two genes is equivalent with the presence of a transcription factor communication between corresponding genes at the level of nucleus. Hormonal communication between cells involves the formation of hormone-receptor complex acting as a transcription factor.

The length of human genes ranges to thousands of nucleotides. This would mean that the longest p-adic length scales of human gene would correspond to $L(173) \sim 16$ micro-meters. The total length of a human chromosome is about 75×10^6 DNA triplets. The corresponding p-adic length scale is $L(193) \sim 2$ cm. The next length scales correspond to the pair $(197, 199)$

and correspond roughly to the size of brain hemisphere and brain. The total length of DNA in chromosomes is $48 \times L(193) \sim 1$ meter, the size scale of human body.

Many-sheeted space-time concept suggests that genes actually correspond to DNA sequences glued to a larger space-time sheet defining the gene. Hox clusters could be one example of this. The geometry of the organism might be coded to these secondary, tertiary, etc. space-time sheet structures of the DNA sequence guaranteeing the coding the topology of the body plan to the topology of the multi-sheeted DNA. These structures are be labelled by p-adic primes and their number would be quite limited.

The linearity of DNA suggests that also the plan of the expression domain should be essentially linear such that each cross section of each module of the expression domain is essentially homogenous phase and its structure is determined by a self organization process constrained by the p-adic length scale hypothesis rather than purely genetically. According to Mae-Wan [I8, I9] living systems are liquid crystals and the basic characteristic of the liquid crystals is that they have crystal like structure in one dimension and are liquids in transversal dimensions [D1] forming. thus layer-like structures. This suggests that p-adic self-organization determines the size of the transversal layer and that DNA only codes some general properties of the liquid phase for a given layer.

The sizes for the expression domains of the genes should form a hierarchy. Effective expression domain can be much larger than the p-adic length scale characterizing it since join along boundaries/flux tube condensates are possible. For instance, the modularization of the genetic programs of plants is perhaps stopped at the level $k = 167$ so that expression domains for plant cells could be regarded as join along boundaries concept of of $k = 167$ plant cells. At the level of organism this perhaps corresponds to the emergence of cell walls hindering the formation of higher level structures formed from cells: plant could perhaps be regarded as a large join along boundaries/flux tube condensate of $k = 167$ plant cells surrounded by a wall. Besides the length of the genome, the number of the p-adic hierarchy levels in the space-time sheet hierarchy of DNA is a natural candidate for a measure of the complexity of the organism.

3.2.2 How the information about morphology is expressed?

One of the fundamental questions of the developmental biology is how the information of genes stored into DNA is translated to the geometry and topology of the organism. The idea of many-sheeted DNA suggests an immediate answer to this question. Expression is "nothing but" the reversal of the coding. The expression domain of the gene contracted effectively to one-dimensional DNA-thread grows back to the expression domain with non-uniqueness and flexibility brought in by self-organization depending on external parameters. This means that various space-time sheets associated with DNA grow during grow to space-time sheets representing actual organs. This process involves the formation of flux tubes between growing space-time sheets associated with various DNA molecules so that coherent macroscopic quantum phases become possible.

One can ask how the growth plan is coded into DNA. Or how much of it is coded into the chemistry of DNA? The idea that DNA is essentially body contracted to a thin thread suggests that the chemical control of DNA is restricted to the local properties of tissues. The space-time sheets of replicating DNA at various body parts simply grow and fuse to form join along boundaries/flux tube condensates growing and giving rise to various organs. The replication of DNA would in turn be quantum self-organization process involving essentially self-hierarchy starting from atomic level and ending up the level of entire organism.

3.2.3 What makes cell differentiation possible?

Cell differentiation is one of the great mysteries of biology. It is known that only part of DNA is active in a cell located in a given part of body and that selective activation of the genome gives rise to differentiation. The problem is to understand the mechanism of activation. Especially difficult challenge for the view about life as mere chemistry is the interaction between large length scales with gene level making possible precise timing of genetic activity.

In TGD framework cell differentiation should correspond to a selection of branch in the the flow diagram describing genetic program. This occurs during the growth since the concentrations of the proteins representing the inputs of the gene programs evolve during the growth and generate also

spatial gradients. Therefore different branches of the genetic program are activated in different parts of the developing organism. Also the genes associated with space-time sheet of increasing size are activated during growth and this brings in new and higher control levels.

Very probably the mechanism involves interaction between microscopic degrees of freedom for DNA and between macroscopic degrees of freedom representing body part where DNA resides. The control and coordination based on Josephson currents between gene space-time sheets and larger space-time sheets is very probably involved as is suggested by the general time scales of genetic activity. Also the interaction with the classical fields of the space-time sheet of the body part to which DNA has wormhole contacts provides an obvious mechanism of activation. The frequencies of the coherent oscillations of em fields involved could be important in both interaction mechanisms. This kind of interactions with larger space-time sheets makes possible to understand induction phenomenon, which corresponds signalling between cells and entire cells groups. This kind of signalling could be crucially important for morphogenesis. Many-sheeted space-time thus provides explanation for the ability of cells to form organs. The notion of cell cohesion is introduced to explain this: the cohesion would correspond to the formation of join along boundaries/flux tube condensate of extended gene space-time sheets.

3.3 What Is The Number Of The Levels In Program Hierarchy?

The obvious idea is that the size of the organism determines the largest p-adic prime contributing to the program hierarchy. It is however not obvious whether to define the size of organism as the "physical", visible size or as electromagnetic size, which is well defined notion in TGD framework.

3.3.1 Does the visible size of the organism determine the number of hierarchy levels?

The simplest working hypothesis is that the number of the levels in the program hierarchy is the number of p-adic length scales between atomic length scale and body size. The larger the visible size of the organism, the larger the number of the levels in the genetic program hierarchy, if this hypothesis is correct. This number is testable characteristic of species and could be valuable guide in attempts to understand how genetic code functions. One can identify the hierarchical level of the gene by looking how many genes it activates before building block protein is coded. It must be however emphasized that visible size need not be a correct criterion: the point is that join along boundaries condensates are possible and give rise to a much larger body size than one might conclude from the value of largest p-adic prime involved.

It is instructive to look the numbers of hierarchy levels in some specific examples assuming that the visible size of the organism determines the number of hierarchy levels. It is assumed that $k = 139$ is the first level which counts as a hierarchy level.

1. Viruses could have 4 hierarchy levels if $k = 139, 149, 151, 157$. Proteins, lipid layers of cell membrane and cell membrane and genes coding building block proteins. It could be that only k=149 is present for the simplest viruses since the formation of the envelope is self-organization process.

2. Bacteria should have 5 levels. $k = 139, 149, 151, 157, 163$.

3. Home fly should have 12 levels since its size is below $L(197) \simeq 1.6$ cm.

4. Animals with size between $L(199) \simeq 16$ cm and $L(211) \simeq 10$ m have 15 hierarchy levels. Note the large gap between $L(199)$ and $L(211) = 64L(199)$.

5. The next level corresponds to the level of dinosauri and whales having sixteen levels unless they correspond to join along boundaries condensates formed from smaller structures which is quite possible. The next level is $L(223)$ and corresponds to size of 640 m!

3.3.2 Does the electromagnetic size of of organism determine the number of hierarchy levels?

There is a large gap between $L(199)$ and $L(211)$ and the next twin length scale corresponds to a length scale of one kilometer. This suggests that new levels of hierarchy possibly emerged

Table 2: Table of p-adic length scales above $L(211) \simeq 10$ meters. $L(151) = 10^{-8}$ meters is assumed.

k	227	229	233	239	241
L_p/m	$2.5E+3$	$5E+3$	$2E+4$	$1.6E+5$	$3.2E+5$
k	251	257	263	269	271
L_p/m	$E+7$	$8E+7$	$6.4E+8$	$5E+9$	$E+10$

after $L(199)$ cannot correspond to the physical growth of body. Mere large size does not guarantee intelligence. Furthermore, if the visible size of the organism determines the number of the hierarchy levels, then dinosauri would have been in a well defined sense more intelligent animals than we! These arguments suggest that the visible size of the organism need not determine the number of genetic program levels.

1. It could be that DNA codes and even controls also the electromagnetic structure of the organism realized as topologically quantized electromagnetic field, "aura", characterizing the organism.

2. An alternative option inspired by the notion of memetic code, which is next level in the hierarchy of genetic codes predicted by the TGD inspired simple model of abstraction process, is that there are higher hierarchy levels present but they are not controlled by the genetic program but call it as a subprogram.

A natural working hypothesis is that EEG correlates with the electromagnetic size of the organism. EEG has emerged rather lately in the evolution and is possessed only by vertebrates. In case of humans it becomes fully developed only at the age of 18. Meditation in general tends to increase the amplitudes of low frequency waves with 8 Hz (alpha wave s) and also waves with lower frequencies (theta wave s). This suggests that growth in electromagnetic degrees of freedom can continue all the lifetime and could be identified as what is called "spiritual growth". It could continue also after the physical death so that the protein based state of life would be only a part of much longer lasting process of self-organization analogous to the development of butterfly. Indeed, in TGD based picture about geometric time the death of the physical body does not mean the end of life.

Schumann resonances are resonances of em fields in the wave cavity defined by the 80 km thick layer between Earth's surface and ionosphere. The frequency range in question correspond to the frequency range of EEG. A hypothesis worth of considering is that human body generates via Schumann resonances topological field quanta, which define electromagnetic sub-selves having the size of Earth. One could even consider the possibility that the highest value of k_G depends on individual and people having tendency to have religious and mystical experiences have exceptionally large value of k_G.

The length scale corresponding to alpha waves is 3.8×10^7 meters and corresponds is roughly 3.75 times the length scale $L(251)$. If levels up to $L(257)$ are present in the human genome then the number of hierarchy levels is 22, not too large number. $L(251) \sim 10^7$ m corresponds to a frequency of 37.5 Hz and is quite near to the 40 Hz frequency claimed by Koch and Crick to be crucial for the visual consciousness! The frequency associated with $k = 257$ corresponds to the frequency of 5 Hz, which also belongs to EEG.

The electromagnetic size of the organ increases rapidly with the number of levels present in the hierarchy as **Table 2** demonstrates.

There are even more explicit observations about the importance of ELF em fields for the functioning of living matter and these observations finally led to a breakthrough in TGD based model of conscious brain. The observations about the special effects of ELF em fields on brain at cyclotron frequencies of ions Na^+, Cl^-, K^+ and Ca^{++} in endogenous magnetic fields $B_{end} = 2B_E/5 = .2$ Gauss were made already at 1983 [J1]. These experiments suggest that these ions/their Cooper pairs form are confined in the magnetic field of Earth and form bound states with macroscopic size of order cell size and with extremely small binding energy corresponding to frequency of order 10 Hz. This is impossible in the standard physics framework but can be understood as resulting from

the dropping of ions and electrons from the atomic space-time sheet to the space-time sheet of the cell where the density of the matter is very low.

In many-sheeted space-time particles topologically condense at all space-time sheets having projection to given region of space-time so that this option makes sense only near the boundaries of space-time sheet of a given system. Also p-adic phase transition increasing the size of the space-time sheet could take place and the liberated energy would correspond to the reduction of zero point kinetic energy. Particles could be transferred from a portion of magnetic flux tube portion to another one with different value of magnetic field and possibly also of Planck constant h_{eff} so that cyclotron energy would be liberated.

Also electron Cooper pairs of high T_c electronic super conductor as well as Cooper pairs of neutrino super conductor are important. Besides magnetic cyclotron frequencies Z^0 magnetic cyclotron frequencies and wormhole cyclotron frequencies make sense: Z^0 currents for ions indeed induce automatically also ionic currents.

One can argue that there is very cold, dry and silent at the cellular space-time sheets and this makes possible macroscopic quantum phases formed by Cooper pairs of ions Na^+, Cl^-, K^+ and electron as well as well as Ca^{++} ions. Later the argument was modified the large values of Planck constant [K5, K3] imply that cyclotron energy scale is above thermal energy at room temperature even if thermal equilibrium of dark space-time sheets with ordinary ones is allowed. Also other ions are possible but these ions are especially important for nerve pulse generation. These super conductors must be effectively one-dimensional (otherwise gap energy is extremely small) and the needed confinement in the transversal degrees of freedom is caused by the presence of B_{end} which could be in TGD framework interpreted as the dark companion of the Earth's magnetic field responsible for controlling biomatter possibly also associated with the personal magnetic body. One could regard these super conductors as associated with the flux quanta of B_{end} having radius 25 μm (the size of a large neuron) by flux quantization and serving as templates for the formation of biostructures.

When the Josephson frequency (potential difference) associated with the weakly coupled super conductors of this kind corresponds to a magnetic transition frequency, quantum jumps between states of either super conductor occur and change the charge distributions and hence potential differences associated with other Josephson junctions associated with either super conductor. Quantum jumps can lead to "wake-up" of either or both super-conductor sub-self giving rise to a mental image. Also emission of ELF photons with resonance frequency is involved. The topological field quanta associated with these photons have typically size of order Earth's circumference. The fact that multiples of the cyclotron frequencies correspond directly to the most important frequencies of EEG and also to some important Schumann frequencies suggests very strongly that the "ELF selves" associated with these topological field quanta represent also selves in our self hierarchy. This leads to a general model for quantum control and for how the space-time sheets representing the self-hierarchy are coupled by join along boundaries bonds serving as Josephson junctions, to a detailed model for the quantum correlates of the sensory qualia and to a model of Boolean mind. ELF selves are a crucial factor of all these models [K6, K7].

The work of Michael Persinger shows that ELF em fields and ELF modulated em fields, affect also gene expression [J3]. Thus it seems that ELF levels, rather than being controlled by gene level, actually control and coordinate gene level rather via the formation of flux tubes between gene space-time sheets and ELF space-time sheet. Whether gene level actually *codes* also ELF levels of the organism is an interesting question. The idea about genome as the entire many - sheeted organism contracted to a thin thread would support this view. On the other hand, the notion of the memetic code identified as the next level of abstraction hierarchy suggests that ELF level corresponds to something genuinely new not reducible to gene level. ELF level could be even seen as a different life form next to the biological life living in symbiosis with biological life. One must also remember that higher levels could couple with gene level only via join along boundaries bonds/flux tubes and that the development of organism could be seen as a "social" process in the sense that growing organism gradually builds flux tubes to the space-time sheets representing higher level selves.

Whether the number of the hierarchy levels in the genetic program hierarchy is larger than the visible size of organism, might be perhaps tested sooner or later by deciphering the number of hierarchy levels in the genetic program. To check the hypothesis about EEG, it is enough to study simplest verterbrates possessing EEG. The identification the levels of various genes in program

hierarchy would mean a tremendous boost in the understanding of genetic code and dramatic change in world view.

3.4 Band Structure Of Chromosomes As An Evidence ForMany-Sheeted DNA?

In prokaryotes DNA is arranged in single chromosome forming closed circular double strand whereas in eukaryotes DNA genome is organized into chromosome pairs. Chromosome is believed to correspond to single DNA thread which has beads in thread structure. Beads are spherical nucleosomes of diameter 10^{-8} meters ($L(151)$!) consisting of histones of 4 different types forming histone octamer. DNA is wound very tightly around nucleosomes, there is about 70 nanometers (slightly less than $L(157)$ of DNA per nucleosome. chromosome forms a helical coil with diameter found to be 30 nm. In interphase chromosomes are coiled once more to a hollow tube of diameter 200 nm (slightly less than $L(167)$) a helix of thickness about 10^{-7} meters. The transition from interphase chromosome to metaphase chromatid is accompanied by a winding to a helical coil of diameter about 600 nm (slightly more than $L(169)$). A possible interpretation of these transformations is as generation of new space-time sheets.

Chromosome banding was discovered already in eighteenth century by Metzner. Chromosome banding characterizes both the chromosome and the method used to produced the banding structure and there are many methods for revealing the band structure. Increasing resolution implies the division of band structures to smaller structures in fractal like manner. The band structures can be divided into two classes.

1. The highly localized heterochromatic bands, nucleolar organizers and kinetochores appear in all organisms. The latter two structures seem to reflect the purely geometrical organization, "packing", of genome rather than the internal organization of genome. The main features of heterochromatic banding are its universality, diversity and variability. Heterochromatic banding is present in all eukaryotes and can differ widely for closely related species and be very similar to widely different species. Heterochromatin seems to correspond to highly repetitive short DNA sequences of 10 nucleotide pairs (10^6 copies) located near the centromere of the chromosome. This DNA is not transcribed to DNA. Pairs are often duplicated and duplication leads to various physiological defects. Soma cells of some organisms appear to have ability to get rid of of heterochromatin whereas it is present in germ cells. These facts suggest that the regions of chromosome near its center regulate gene expression and that highly repetitive DNA sequences represent sites for genes at which repressor proteins bind. Abnormally large duplication of repressor sites would lead to stronger repression is more effective and could lead to abnormal development.

2. The chromosomes of the eukaryotes contain also non-localized bands called euchromatic bands. Patterns of euchromatic bands resemble closely to the patters of DNA replication and patterns correlate very strongly with species. Thus euchromatic bands correspond to active RNA. The moderately repetitive DNA which is transcribed corresponds corresponds to euchromatin. It is known that there are several types of euchromatic banding. Band patterns can be used as diagnostic tools to identify various chromosome fusions and splittings. Various bands are of enormous value in providing manner to locate genes in genome.

In TGD framework a natural interpretation of various types euchromatic banding provide evidence for the many-sheeted DNA. Thus euchromatic banding should reflect the modular structure of the genetic program as well as the interspersing of control regions and transcribed regions of genes corresponding to the basic structure "If A then B" of the gene.

4 About The Notion Of Genetic Program

This section is devoted to a general discussion of several key ideas which were not available when I formulated the first TGD inspired vision about genetic programs and morphogenesis. The recent view about TGD inspired theory of consciousness and quantum biology allows much deeper formulation for the notion of genetic program. In particular, the vision about DNA as topological quantum computer allows to identify a possible role for the intronic portions of DNA.

4.1 What The Notion Of Genetic Program Could Mean?

How to define the notion of genetic program? As TGD inspired view about consciousness and living systems has evolved several several alternative definitions have emerged.

1. The first attempt to answer the question was inspired by the observation that so called Combinatorial Hierarchy might define a hierarchy of codes containing at least genetic code and what I christened as memetic code. The definition is based on the interpretation of DNA sequences - at least genes - as Boolean statements, which define axioms of a formal system which are identically true. The system in question corresponds to Mersenne prime M_7 containing 127 statements about which 64 is identically true and define the axions in question. Memetic code would correspond o M_{126} with 2^{126} axioms and could be realized as sequences of 21 DNA codons.

   The first proposed realization of the memetic code was in terms of temporal sequences of what I called cognitive neutrino pairs. It looks too nonsensical in the framework of standard physics and also in TGD framework it looks questionable.

2. With the advent of zero energy ontology (ZEO) came the observation that one could code "laws of physics" by zero energy states for which positive and negative energy parts of the state correspond to initial and final states for time evolution which conserves total fermion number and other charges. This would give rise to pairs of statements which are consistent with each other identifiable as rules expressed as superpositions of instances which are consistent with the rule.

3. DNA as topological quantum computer vision [K4] led to a considerable progress in the understanding of what genetic programs might mean. The basic idea is that the braiding of flux tubes connecting DNA nucleotides/codons to lipids of nuclear/cell membrane define natural topological quantum computation programs. It is natural to assign them with the intronic portion of the genome. These programs would define the software of the living matter whereas genes would define the hardware. The approach suggests a mapping of the letters A, T, C, G of the genetic code to u and quarks and their antiquarks. The discovery of what I call dark DNA realized as states of dark protons led to alternative realization of the sequences of DNA codons.

4. A deeper understanding of the anatomy of quantum jump in zero energy ontology led to a further insights [K19, K17]. In this framework the state function reductions to the opposite boundaries of causal diamond (CD) produce zero energy states with opposite inherent arrows of imbedding space geometric time. These two reductions occur alternately and have interpretation in terms of sensory perception followed by motor action. This process occurs in all length and time scales and is not a property of only biosystems. Memories are represented as negentropically entangled states assignable to CDs and by Negentropy Maximization Principle [K10] they are approximate invariants and defined what might be called self representations or representation for the world. The braiding of the magnetic flux tubes is space-time correlate for the negentropic entanglement.

   The important outcome is that time reversal of memory representation can be interpreted as a plan or prediction realized in terms of braidings of the flux tubes. The interpretation of genetic program would be naturally in terms of the time reversed memories determining how, say, the seed of plant evolves to a plant. Dark photons are key element in the reading process and they can decay to ordinary photons with same energy identifiable as bio-photons. Biophotons give hopes of proving experimentally the existence of the template of the full-grown organism defined by its magnetic body and there is evidence that this kind of template indeed exists.

In the following the earliest proposals for what genetic programs might be, are discussed. These proposals can quite well be consistent with more recent views but are more formal and lack direct connection with the recent view about TGD inspired theory of consciousness and of living matter.

4.2 Genes And Genetic Programs

A possible model for the genetic program is based on the interpretation of DNA sequences as statements of a formal system. 64 basic statements represented by DNA triplets correspond to axioms which are identically true. Gene G is interpreted as a theorem of type "If I_1&....&I_n true then O_1&O_2...&O_n true" or generalizations thereof obtained by adding several input statements and allowing also negations of the input statements. It could be that complementary DNA strand represents statements of type "If I_1&....&I_n true then O_1&O_2...&O_n not true". If given DNA sequence has been translated to give rise to some minimum concentration of protein coded by it, the truthness of this statement is emotionally experienced by the genetic computer. Otherwise the truth value of the statement is ill defined.

The information about truth value of I_k is represented by the catalytic action of the enzyme coded by I_k on those parts of gene which are not transcribed, in particular on the promotor sequence of the gene. They give rise to experiences about the truth values of the statements $I_1...I_n$. The conclusion "$O_1\&,..,\&O_n$ is true" represents the output of the gene. These statements appear as the premises for the statements represented by some other genes. Thus the running of genetic program represents a sequence of becoming conscious about various kinds of truths of the formal systems represented by genome.

4.2.1 Genes as statements of conscious formal system

The assumptions for the model of genes as statements of a conscious formal system are following.

1. DNA sequences are assumed to represent axioms of some formal system. There are 64 basic triplets of DNA, which correspond to basic independent statements, axioms, from which higher level statements are built using many-sheeted DNA making possible the construction of statements about statements about....

2. Genes are assumed to represent statements of type "If I_1&...&I_m then O_1&...&O_n true." or "If I_1&...&I_m then O_1&...&O_n not true." depending on which strand of DNA double helix is transcribed.

3. The output of gene is either true or ill defined gene can only be conscious about truthness of a statement or be unconscious. The logic of our conscious experience (which need not have anything to do with the possible gene level mind) is consistent with this. We can have conscious experience about truth values of a very limited set of statements and have no experience about the infinitely of all possible logical statements. Note that the experience with standard logic would suggest that the value of the statement is either true or false. The experience with ordinary computers would in turn suggest that the value of outputs could be true, false or ill defined.

4. Gene expression is the counterpart for the emotional experience about truthness of the statement represented by the gene. One can say that the output of the gene is true or ill-defined. The output of the gene G generated by gene expression appears as an input for those genes G_i, which represent statements whose premises depend on the statements appearing in the output of G. This implies that the evolution of the genome involves a sequence of conscious emotional experiences about truthness of some statements, which in turn make possible to become conscious about truthness of some other statements.

5. The formal system represented by DNA and many-sheeted DNA sequences has certain Gödelian flavor in it. The truthness of the stopping sign DNA triplets UAA, UAG and UGA cannot be expressed in terms of proteins coded by exons. They are like undecidable statements in an axiomatic systems whose truth value cannot be deduced from the axions. Introns however can contain stopping sign and if mRNA-protein complexes represent truthness of intronic statements, then it is possible to experience the truthness of also the statements containing stopping sign DNA. The analogy with "holy trinity" of mystic and religious thinking is obvious: whether "holy trinity" exists is not decidable by human means! Proteins clearly have finite expressive power. Given amino-acid coded by DNA triplets X_i tells only that the statement $\vee_i X_i$ is true so that the number of emotionally experiencable basic truths reduces to 20.

A more precise but less general model for genome as a conscious formal system is based on the following assumptions.

1. Exon part G of gene and its intron parts I_i are interpreted as logical statements G and I_i. Together they represent the conclusion of a theorem. Hence the running of genetic program means generation of a sequence of conscious experiences of type "...*then* $G_1\&...\&G_n$ is true" or "...*then* $G_1\&\cdots\&G_n$" is false depending on which DNA strand is involved in transcription.

2. The statement represented by a gene G is experienced to be true if G has generated some minimum concentration of the corresponding protein $P(G)$. Under what conditions the intronic statements $I_1,\cdots I_n$ are experienced as being true, is not clear. The basic dogma of genetics suggests that also now overcritical values for the proteins $P(I_i)$ translated from intronic mRNA inside nucleus make possible conscious experience about the truthness of these statements.

3. mRNA-protein complexes are not able to regulate directly the activity of genes having intron statements as premises. The regulation mechanism, if present, must be indirect. Simplest regulation mechanism is based on the observation that gene can be in states in which the roles of some introns and exons are changed. In particular, each gene has complementary gene for which exons and introns (with stopping sign excluded) have changed their roles. If the protein coded by complementary gene has the role of a silencer for the expression of gene protein, the production of mRNA-protein complexes means that the translation of the silencer molecules does not occur anymore. Since the existing silencer molecules gradually decay, these genes are activated. Thus the activation of a gene by intron-exon transformation activates automatically also the genes whose input depends on the output of the gene. Note that exons would code what is very much analogous to "printed output".

It is important to notice that cognition represented basically by a genetic program is not restricted to brain. All cells of body have DNA and elementary cognitive abilities. Perhaps the special role of the frontal lobes in cognition is due to the fact that DNA in cortex has wormhole contacts with topological field quanta representing electromagnetic part of our body not visible to bare eyes. The electromagnetic size of our body could be given the size of entire Earth as suggested by the fact that EEG frequencies correspond to Schumann frequencies associated with the resonances of the wave cavity between Earth's surface and ionosphere.

4.2.2 Genes as modules of a genetic program

Genes can be also regarded as modules of genetic program. The input of module consists of the premises of the statements $I_1\cdots I_n$ and conclusions represent the statements "$O_1\cdots ...O_n$ true" or "$O_1\cdots ...O_n$ not true". $O_1\cdots O_n$ serves as input for other genes and running of genetic program continues as long as premises of some genes are true.

Ordinary programs "If A then B" and "While A do B" as basic control structures. Also genetic program should have some control structures and it seems that both these structures appear in genetic programs.

1. Assuming that genes represent theorems, the general form of statements appearing in genetic program is indeed "If A then B". A is represented in terms of enzyme concentrations and possibly intron mRNA-protein complexes activating or repressing some genes. B is represented by the protein and intronic mRNA-protein complexes transcribed from gene.

   "While $N>0$ do $\cdots$ and $N\to N-1$" is the basic structure of loop in computer program. The simplest function of this loop is to circumvent infinite loops, which are nightmare of programmer. The counterpart of an infinite loop is member of species, which never dies. The presence of too many sufficiently intelligent (and, as one could bet, selfish) individuals of this kind would be a catastrophe from the point of view of species. It seems that this kind of condition is represented. Certain genes crucial for the survival of higher organisms contain certain amount of telomere which is reduced every time when the gene is active. This means that the gene modules effectively contain additional condition "While the amount of telomere $>0\cdots$ and reduce the amount of telomere by one unit".

One can understand basic rules of Mendel easily if the conditions guaranteeing the activity of gene depend on both members of chromosome. For instance, in the case that two traits are such that second on dominates, the output of gene realizing the property is logical AND function having as its arguments the truth values of the statements represented by related genes in the chromosomes. If G_1 in chromosome 1 and G_2 in its partner chromosome have coded sufficient amounts of their specific enzymes then some third G codes the enzyme giving rise to the non-dominant trait, otherwise the dominate trait appears. Second situation is the one in which traits are not yes/no properties but have discrete spectrum like red, white and pink. In this case there are two genes G_1 and G_2 coding red and white directly. Various combinations of color would correspond to three combinations G_1G_1 (red), G_2G_2 (white) and G_1G_2, G_2G_1 (pink).

The notion of many-sheeted DNA implies that gene submodules should correspond to body parts. This raises several interesting questions about what the bodily correlates of the "body part =subprogram calls second body part=subprogram". Is there some kind of communication occurring between body parts also? Perhaps chemical communication based on the same enzymes that regulate gene transcription? Gene modules call gene modules corresponding to smaller sized body parts: is this true also for the communication between body parts? Or does signal transfer occur in dual manner from small body parts to larger ones?

If this kind of correspondence exists, say, at hormonal level, one could directly deduce information about the structure of genetic program from the topology of the hormonal communications. For instance, genome should have central unit analogous to brain receiving information from controlling the activities of the rest of the genome. An obvious candidate for the brain of chromosome is centromere. Centromere seems to be analogous to brain in the sense that lot of repetitive DNA presumably controlling gene expression is situated near the centromere.

4.2.3 What about cognitive fermion pairs as representation of memetic codons?

According to the original speculations, memetic code could be represented as temporal sequences of what I called cognitive neutrino pairs with neutrino and antineutrino assigned with the opposite throats of a wormhole contact carrying Z^0 magnetic flux and assumed to be associated with cell membrane. The possibility of these sequences was argued to be allowed by the failure of the strict determinism for Kähler action. This idea might indeed make sense if cell membrane is near vacuum extremal and corresponds to a scaled variant of weak interaction physics with the weak length scale being of order cell size scale so that classical Z^0 and W fields become important (both non-standard value of $\hbar_{eff}$ and different p-adic length scale for weak bosons could give rise to scaled variants of weak interaction physics). From the point of view of standard model the idea is of course complete non-sense. The basic support for the idea about scaled variants of weak interaction physics is that it could understand the mysterious large parity breaking in living matter in this manner.

The weak point of the original idea was the assumption that the negative sign for neutrino energy is due to its interaction with classical Z^0 electric field. The idea however generalizes considerably in the framework of zero energy ontology (ZEO) inspired much later by the non-determinism of Kähler action. Causal diamond (CD) becomes the basic geometric objects and is identified as correlate for a spotlight of consciousness. The simplest zero energy states assignable to CD would correspond to fermions with opposite quantum numbers localized at the two light-like boundaries of CD (this only approximately since second boundary correspond to a superposition of states appearing as final states for the possible decays of the fermion).

The positive energy fermion has fermion number, weak isospin, and spin and allows representations of binary digit. In particular, nucleotides A, T, C, G could be represented in terms of these quantum numbers. The representation of in terms of u, d and their antiquarks is only one representations. The idea is that one can construct negentropically entangled temporal sequences of this kind of CDs having interpretation as bit sequences. This could give rise to a representation of also memetic codons as sequences of 126 binary digits represented in terms of fermions or as sequences of 21 triplets of binary 2-digits representable as zero energy states representing dark DNA codons.

4.2.4 Could DNA level contribute to our consciousness

The fermionic Boolean algebra defined by fermionic Fock state basis is a natural candidate for a quantum physical correlate for Boolean thoughts [K7]. In zero energy ontology (ZEO) the pairs of positive and negative energy fermion states with opposite net fermion numbers are in turn natural candidates for pairs of Boolean statements consistent with each other: basically they could code for the "laws of physics".

This suggest that some part of neuronal genome, whose dynamics is slow as compared to neural dynamics but consistent with the slow dynamics of the belief system as required, stores information, not only about the physical structure of organism, but also about our belief structures, intentions and long term goals. These beliefs become conscious when gene is activated, perhaps by neural activity: neural transmission is indeed known to induce effects at gene level. The statements represented by genes would correspond directly to our conscious beliefs at various levels of the space-time sheet hierarchy! Perhaps the memetic codewords consisting of 21 DNA codons (also dark DNA codons) could have interpretation as bit sequences representing integers in the range $(1, 2^{126})$ and genuine Boolean thought resides basically at the genome level.

This mechanism becomes possible only at p-adic length scales longer than neuron size scale. Therefore the emergence of genes with a total length larger than the p-adic length scale $L(169) \simeq 5$ μm should correspond to the emergence of the eukaryotes, introns and Boolean consciousness. Thus only some large enough bacteria (at least Fischerella) can have cognitive abilities. The absence of mitochondrial introns is probably due to the short length of the mitochondrial genes. The mitochondrial DNA of fungi is known to contain introns and the reason is perhaps that the length of DNA is longer than the critical length. Obviously, the proposed scenario implies a relationship between DNA level and logical thinking and partial reduction of language structures to DNA level.

In this picture genome is a collection of genes which can be regarded as Boolean algebras $B(N)$ defined by N independent binary digits. Prokaryotes would represent only $B(1)$ algebra and only single statement would be associated with gene: the only possible conscious experience associated with gene statement would be experience about the truthness of the corresponding statements. Higher level Boolean mind realized as Boolean algebras $B(N)$, $N > 1$ are present only in eukaryotes since only their DNA can have Z^0 wormhole contacts to space-times sheets $k \geq 169$. Also the genes associated with cell organelles corresponds almost as a rule to the smallest possible Boolean algebra.

4.3 DNA As A Topological Quantum Computer

For years ago I developed a model of topological quantum computation (TQC) combining TGD based view about space-time with basic ideas about topological quantum computation and ended up with the proposal that DNA might act as a topological quantum computer. One can imagine several manners in which DNA or RNA could act as a topological quantum computer and it good to try to state clearly what one wants.

1. Natural requirements are that the topological quantum computer programs can be naturally combined to larger programs and evolution means this kind of process; that the programs have a natural modular structure inherited from the previous stages of evolution; and that the computation is not restricted inside single nucleus.

2. DNA and/or RNA defines the hardware of topological computation and at least for more advanced topological quantum computers this hardware should be static so that only programs would be dynamical. This leaves only DNA in consideration and the entangled initial and quantum states at the ends of braids quantum states would be assignable to static DNA structures.

3. The program would be determined by different braidings connecting the states of DNA in time direction or in spatial direction. Since the genomes are identical in different nuclei, the strands could connect different nuclei or conjugate strands of double DNA strand. Reconnection process would allow to modify the hardware for TQC.

4.3.1 The recent progress in quantum TGD and TGD inspired quantum biology

After the advent of the first model for topological quantum computation in TGD Universe [K14], the mathematical and physical understanding of TGD has developed dramatically and the earlier quite speculative picture can be replaced with a framework which leads to a rather unique view about topological quantum computations by DNA.

1. Universe as a topological quantum computer

One can say that the recent formulation of quantum TGD states that the entire Universe behaves like a topological quantum computer. This notion of topological quantum computer differs however from the standard one in many respects.

1. The emergence of hierarchy of Planck constants realized as a generalization of the notion of imbedding space is now a basic piece of TGD allowing an elegant formulation of quantum TGD [K15, K5]. The phases of matter with large Planck constant are interpreted as dark matter. Large values of Planck constant make possible topological quantum computations in arbitrary long time scales so that the most fundamental objection against quantum computation can be circumvented.

2. Zero energy ontology forces to unify S-matrix and density matrix to M-matrix - the product of the square root of density matrix and S-matrix- defined as time-like (or rather light-like) entanglement coefficients between positive and negative energy parts of zero energy state [K2, K1]. Connes tensor product emerging naturally from the notion of finite measurement resolution described in terms of inclusions of hyperfinite factors of type II_1 defines highly uniquely the M-matrix. M-matrix would be natural candidate for defining topological quantum computation in light-like direction. Connes tensor product makes sense also in space-like direction and would define quantum storage of functions represented as entanglement coefficients.

3. The notion of number theoretic braid [K16, K2] has become a basic element of the formulation of quantum TGD based on the requirement of number theoretical universality. As a matter fact, the notion of braid is generalized in the sense that braid strands can fuse and decay. One can make guesses about the details of the braid motion, which do not however matter since only topology is what matters. One could argue that the braid orbits must be light-like curves: the generic curve is locally space-like. The motion of braids corresponds to boundaries of string world sheets and the vanishing of induced weak fields at them gives a strong constraint on them. Note that one expects Kac-Moody type invariance deforming the light-like 3-surfaces and respecting their light-likeness but without changing the Kähler action and conserved quantities associated with preferred extrema.

   For generalized Feynman diagrams partonic light-like 3-surfaces meet at 2-dimensional vertices defined by partonic 2-surfaces [K1]. This implies that braids replicate at vertices: the interpretation is as a copying of classical information. Quantum information is not copied faithfully. The exchange of partonic 2-surfaces in turn corresponds to quantum communications. Hence quantum communication and quantum copying emerge naturally as additional elements. Space-like Connes tensor product in turn defines quantum memory storage.

4. Computation time is a fundamental restriction in both ordinary and quantum computation. Zero energy ontology makes possible communications in both directions of geometric time, which suggests the possibility of geometric time loops in topological quantum computations. Could this mean that computation time ceases to be a restriction and ordinary computations lasting for infinite amount of geometric time could be performed in a finite time interval of observer's time? This is perhaps too much to hope. The subjective time taken by the computation would be infinite if each step in the iteration corresponds to single quantum jump. If this is the case and if each quantum jump of observer corresponds to a finite increment of geometric time perceived by the observer, time loops would not allow miracles.

2. The notion of magnetic body and the generalization of the notion of genome

The evolution of ideas related to quantum biology provides also new valuable insights. In particular, the notion of magnetic body leads to a model of living system in which dark matter at magnetic flux quanta of the field body of biological system uses biological body as a motor instrument and sensory receptor [K3]. Quantum control would be naturally via the genome and sensory input would be from cell membrane containing all kinds of receptors. This would suggest that magnetic flux sheets traverse through DNA strands and cell membranes.

The quantization of magnetic flux with unit defined by Planck constant having arbitrarily large values leads naturally to the notions of super-genome and hyper-genome [K9]. Super-genome would consists of DNA strands of separate nuclei belonging to single magnetic flux sheet and these sequences of genomes would be like lines of text at the page of book. Super-genomes in turn can combine to form text lines at the pages of a bigger book, I have used the term hyper-genome. This hierarchy of genomes would give rise to a collective gene expression at the level of organs, individuals of a species, and at the collective level consisting of populations containing several species. Even biosphere could express itself coherently via all the genomes of the bio-sphere. The model of topological quantum computation performed by DNA should be consistent with this general picture.

4.3.2 Model for DNA based topological quantum computation

The most promising model of DNA as topological quantum computer relies on the hierarchy of genomes. The flux sheets or collections of parallel flux tubes assignable to a magnetic body would traverse the DNA strands of several nuclei so that strands would be analogous to lines of text on the page of a book.

DNA strands would define the intersections of magnetic or number theoretic braids with plane and braiding would be associated with with the magnetic field lines or flux tubes transversal to DNA. The M-matrix defining topological quantum computation would act on quantum states assignable to nucleotides.

1. The interpretation of nucleotides

The interpretation of the A, T, C, G degree of freedom is not obvious and one can consider several options.

1. The quantum numbers entangled by braids having nothing to do with (A, T, C, G) assignable to nucleotides and the braiding does not affect nucleotides.

2. The nucleotides (A, T, C, G) correspond to four different colors (a, t, c, g) for braid strands with conjugate nucleotides defining conjugate colors. The subgroup of allowed braidings would preserve the color patterns. The minimal assumption consistent with the mapping of nucleotides to quarks and antiquarks [K4] is that braid strands connect only nucleotides and conjugate nucleotides.

3. The model requires that the genomes in different nuclei are identical: otherwise it is not possible to realize braidings as symmetry transformations mapping portions of DNA to their conjugates (as noticed, this map would not occur at the chemical level). An interesting question is whether also the permutations of nucleotides of different codons are allowed or whether only codons are permuted so that they would define fundamental sub-programs.

4. One can understand why the minimum number of nucleotides in a codon is three. The point is that braid group is non-commutative only when the number of strands is larger than 2. The braidings acting as symmetries would correspond to a subgroup of ordinary braidings leaving the color pattern of braid invariant. Obviously the group is generated by some minimal number of combinations of ordinary braid generators. For instance, for two braid strands with different colors the generator is e_1^2 rather than e_1 (two exchange operations/full 2π twist). For codons one would have four different subgroups of full braid group corresponding to codons of type XXX, XYY, XXY, and XYZ. Each gene would be characterized by its own subgroup of braid group and thus by an M-matrix defining topological quantum computation.

5. It might be possible to understand the "junk DNA" character of introns. Introns are the most natural candidates for the portions of genome participating topological quantum computations. The transcription process would disturb topological quantum computation so that

introns should be chemically passive. Since the portion of "junk DNA" increases with the evolutionary level of the species evolution would indeed correspond to an increase the amount of topological quantum computations performed.

2. Two realizations of topological quantum computation and their combination

One can imagine two basic realizations of topological quantum computation like processes- or to be more precise - entanglement by braiding. In TGD framework this entanglement could be interpreted in terms of Connes tensor product.

2.1 Space-like entanglement

The first realization would rely space-like braids. Braid strands would connect identical lines of text at the page of book defined by sequences of genomes of different nuclei. Inside nucleus the strands would connect DNA and its conjugate. The braiding operation would take place between lines.

In this case it would be perhaps more appropriate to speak about quantum memory storage of a function realized as entanglement. These functions could represent various rules about the behavior of and survival in the physical world. For this option A, T, C, G cannot correspond to entangled quantum numbers and the interpretation as braid colors is natural. Braiding cannot correspond to a physical braiding of nucleotides so that (A, T, C, G) could correspond to braid color (strands would connect only identical nucleotides).

Strands would not connect strand and its conjugate like hydrogen bonds do but would be like long flux lines of dipole field starting from nucleotide and ending to its conjugate so that braiding would emerge naturally. Color magnetic flux tube structures of almost atom size appear in the TGD based model of nucleus and have light quarks and anti-quarks at their ends [L1] , [L1]. This could be the case also now since quarks and anti-quarks appear also in the model of high T_c superconductivity which should be present also in living matter [K3].

2.2. Light-like entanglement

Second realization would rely on light-like braids at the boundaries of light-like 3-surfaces connecting 2-surfaces assignable to single genome at different moments of time. Braiding would be dynamical and dance metaphor would apply. The light-like surface could intersect genomes only at initial and final moments and strands would connect only identical nucleotides. Light-likeness in the induced metric of course allows the partonic 3-surface to look static at the level of imbedding space.

What number theoretic braid is, is far from clear. I have proposed the identification of fundamental number theoretic braids in terms of the minima of the Higgs like field associated with the Kähler-Dirac operator would be very natural in this case. An alternative definition of braids would as boundaries of string world sheets at which spinor modes are located. This definition is not however completely unique.

Genes would define only the hardware unless they code for the magnetic body of DNA too, which looks implausible. The presence of quantum memory and quantum programs would mean a breakdown of genetic determinism since the braidings representing memories and programs would develop quantum jump by quantum jump and distinguish between individuals with the same genome. Also the personal development of individual would take place at this level. It would be these programs (that is magnetic bodies) which would differentiate between us and our cousins with almost identical genome.

2.3 Combination of the two realizations

These two variants of TQC accompany each other automatically if DNA nucleotides are connected to the lipids by magnetic flux tubes [K4]. In this case the 2-D flow of lipids induced by the self organization pattern of the metabolically induced flow of cellular water would induce the TQC as dance and this in turn would would generate braiding of flux tubes connecting lipids to the nucleotides. Presumably a gel-sol transition of cytoplasm accompanies TQC in this kind of situation.

4.3.3 Biological evolution as an evolution of topological quantum computation

This framework allows to understand biological evolution as an evolution of topological quantum computation like processes in which already existing programs become building blocks of more complex programs.

1. The transition from RNA era to DNA era involving also the emergence of cell membrane bounded structures would mean the emergence of the topological quantum computation using a static hardware.

2. For mono-cellulars double DNA strands define space-like topological quantum computations involving only single step if the braids connect the nucleotides of the two DNA strands: obviously a reason why for double DNA strands.

3. For multicellular organisms more complex space-like topological quantum computations would emerge and could code rules about environment and multicellular survival in it. At this step also introns specialized to topological quantum computation would emerge.

4. A further evolution as a generation of super-genomes in turn forming hyper-genomes and even higher structures would have a concrete counterpart as the organization of braids of lower level to form braids at higher level so that topological quantum computations would become increasingly complex and program module structure would emerge very naturally.

5 Ideas About Concrete Realization Of Genetic Programs

In this section some ideas about concrete realization of genetic programs are discussed. The ideas derive from the first years of TGD inspired quantum biology and it would be interesting to see how the systematic application of recent understanding of TGD could enrichen the picture.

5.1 How Gene Expression Is Regulated?

In case of prokaryotes the regulation of transcription is quite satisfactorily understood. The problem is to understand how transcription is regulated in case of the eukaryotes and here the notion of many-sheeted DNA could be crucial.

5.1.1 Operon theory for the regulation of gene expression in prokaryotes

Jacob, Monod and Pardee [I3] suggested operon theory for the regulation of the transcription of genes responsible for lactose production in E. coli. The presence of lactose induces E. coli to produce 3 enzymes needed in the production of lactose. The enzymes correspond to three structure genes x, y, z of lactose. The mechanism is following.

1. So called i gene regulates the production of the building block proteins of the repressor protein which self-assembles as tetramer of the repressor protein.

2. Repressor protein binds to a specific site next to the promotor and hinders the binding of the RNA polymerase so that transcription becomes impossible.

3. Inductor, in the present case lactose, binds to the repressor protein and hence hinders the formation of the repressor tetrameres so that transcription becomes possible.

4. The transcript contains not only the gene but also the entire operon so that several genes are translated simultaneously.

5.1.2 How eukaryotes differ from prokaryotes?

Operon theory does not generalize as such to eukaryotes. Although the notion of the promoter generalizes, there is no clear-cut evidence for operons [I3]. Rather, silencers and enhancers could take the role of the inducers and repressors in the eukaryotic gene expression. The action of a silencer/enhancer is not sensitive to its precise location or orientation and the distance from

promoter can be more than thousand nucleotide pairs. TGD based explanation is based on the notions of many-sheeted DNA and protein. Silencers and enhancers mediate the interaction of the atomic space-time sheet of DNA with the classical fields of some larger space-time sheet. This interaction makes possible top-bottom type control analogous to the control of slave by master in Haken's theory of self-organization. Both classical em and Z^0 fields can control the gene expression in this manner. In eukaryotes classical Z^0 fields could have especially important role. Classical Z^0 fields are believed to be crucial for the model of cognition and this suggests that "mind-matter" interaction could at least partially relay on classical Z^0 fields and enhancers/silencers. Silencers and enhancers could make also possible Josephson junctions between gene space-time sheet and some larger space-time sheet and thus realize "biofeedback".

Second difference is related to introns. It is known that introns and exons can change their roles and it is known that there are several pathways for splicing leading to different proteins, isoforms. The replacement of single-valued gene→ polypeptide map with many-valued map obviously increases the information content of gene. The interpretation of introns and exons as two interspersed computer codes with intron lines of code separated from exons by "comment signs" marking each nucleotide of intron is attractive model for the situation. Dropping of some comment signs changes the result of the splicing process. Comment sign distribution could be dynamical and tissue specific so that one could say that genome is not invariant of species but only of a particular tissue type. This obviously reduces the genetic determinism and gives organism better abilities to survive. In human genome 1 percent of gene corresponds to exons in the "normal" state (in whatever manner that state is defined). The number of various combinations of exon and intron combinations is 2^N where N denotes the numbers of basic components of gene (perhaps coding proteins having no decomposition to modular proteins). The number of combinations increases exponentially with N and provides huge flexibility.

The interpretation of various exon-intron decompositions as statements of Boolean algebra of statements about N basic statements suggests strongly itself. In particular, exonic and intronic proteins for same intron-exon decomposition would naturally correspond to a statement and its negation. Thus one could regard eukaryotes as representing higher levels in the hierarchy of abstractions in which prokaryotes represent the lowest level. Eukaryotic genome would be a collection of genes identifiable as Boolean algebras and the running of the genetic program would mean that at given moment some statement in some of these Boolean algebras is experienced to be true. This kind of identification would mean effectively understanding of the logical meaning of the genetic code. The dominance of particular exon-intron decomposition over its complement would simply mean that this Boolean statement is true while its complement is not true. This suggests the possibility that only those 2^{N-1} exon-intron configuration which represent statements consistent with a given atomic statement are usually realized (atomic statement corresponds to a one-element subset in the set theoretic representation of Boolean algebra).

The fact that introns and exons are set theoretical complements of each other raises the possibility that the proteins coded by introns and exons have opposite effects as transcription factors. For instance, exons could code enhancer and introns could code silencer. This implies that same gene can act as both enhancer and silencer. The only thing needed is that the roles of introns and exons are changed. When this occurs, the production of intron-RNA complexes begins and the production of the composite protein coded by introns acting as exons stops. When the production of silencer protein coded by introns ceases, the silencer proteins associated with the operator sites gradually decay and gene expression also enhanced by enhancer proteins can start. Thus the activation of the gene module activates automatically the gene modules which it calls and genetic program runs.

This would suggests rather general mechanism of gene expression.

1. There exists 2^{N-1} exon-intron decomposition plus their complements. Depending on the state of gene with given exon-intron decomposition, exons or introns are translated to protein. Both introns and exons represent statements. The modular decomposition of protein to subproteins often represented by the decomposition to exons corresponds to the decomposition of the statement E to $E = E_1\&E_2...\&E_n$. Same holds true for I: $I = I_1\&I_2...\&I_n$.

2. The proteins coded usually by exons activate some genes and these proteins appear as prerequisites of type

$$IF\ [E_1\ OR\ E_2 \cdots\ OR\ E_n]\ THEN \cdots\ .$$

The presence of all activators is not necessary.

3. Introns correspond to repressors quite generally. The proteins coded by introns in the "abnormal" state of the gene correspond to prerequisites of type

$$\cdots IF\,[NOT(I_1)\&NOT(I_2)...\&NOT(I_n)]\ THEN \cdots\ .$$

This means that the absence of all repressors from operator site is necessary. The general form of gene statement is

$$IF\ [E_1\ OR\ E_2 \cdots\ OR\ E_n]\,\&IF\,[NOT(I_1)\&NOT(I_2)...\&NOT(I_n)]\ THEN \cdots\ .$$

Various E_k: s and I_k: s represent proteins in turn having modular decomposition to a product of more primitive statements.

5.1.3 The role of the hierarchy of Josephson currents

The control– and coordination hierarchy formed by super conductors represented by space-time sheets coupled to each other by join along boundaries bonds suggests new quantum level control mechanisms for genetic expression. Josephson currents at resonant frequencies corresponding to some magnetic transition of gene or its substructure could "wake-up" the gene self and initiate the self-organization process leading to the gene activity. Silencers and enhancers could correspond to proteins which have join along boundaries bonds/flux tube contacts to larger space-time sheets serving as masters. This would explain why neither the exact position nor orientation of the silencer or enhancer is not important for their functioning.

Various genes and and associated control structures have mutual Josephson junctions controlling gene expression. This would mean the presence of extremely weak longitudinal electromagnetic fields (the potential differences over Josephson junctions would be in 10^{-14} eV range). The control mechanisms behind morphogenesis are poorly understood and phase gradients along chromosomes and along the growing organism could be involved with the control of morphogenesis. The fact that the rate of the translation process is about 20 amino-acids per seconds is in accordance with the idea that this process is controlled by a Josephson current associated with some ion having this frequency.

5.2 Model For The Physical Distinction Between Exons And Introns

Introns seem to begin with nucleotide pair CT and end with pair TC: the assumption that this is the sole signature of the beginning of intron is however not consistent with the observation about the change of the roles of intron and exons. Furthermore, this criterion is obviously not sufficient for telling where intron begins and ends since also ordinary genes can contain similar section.

There are several constraints on the marking telling whether a given nucleotide is exon or intron.

1. The splicing mechanism or DNA telling which portions of it are transcribed to mRNA should rely on the marking mechanism. This requires that the enzymes responsible for the splicing are able to distinguish between introns and exons.

2. Introns can start and end in the middle of DNA triplet. This suggests that the marking is assigned with each nucleotide separately rather than each codon.

3. Both intronic and exon portions of DNA are transcribed to mRNA and splicing of the intronic portions of mRNA occurs only after that. Therefore the total length of the exon portions of DNA corresponding to a given gene must correspond to integer multiple of codons and there must exist some mechanism forcing this. The length of the intronic portions seems to be free.

4. The transformations of exons and introns to each other are known to occur. Tere must exist a physical mechanism changing the marking.

5.2.1 Could magnetic flux tubes serving as braid strands distinguish between introns and exons?

A priori one can imagine an endless number of possibilities for the physical realization of the binary digit distinguishing between exon and intron nucleotide. The vision about DNA as topological quantum computer with flux tubes connecting introns to the lipids of nuclear and cell membranes (not necessarily those of same nucleus or cell) acting as braid strands would realize this distinction naturally: the introduction of the connecting flux tube would transform exon an intron. Reconnection of pairs of flux tubes between DNA nucleotides or codons and lipids would allow to effectively eliminate the braid strands and transform pair of introns to exons. The first guess would be that the presence of braid strand could make it impossible to perform transcription. This is not the case: both intronic and exon portions are transcribed to mRNA and slicing process cuts off the intronic portions of mRNA after that.

1. The original model of DNA as topological quantum computer [K4] assigns to the nucleotides A, T, C, G quarks u and d and antiquarks $\overline{u}$ and $\overline{d}$. This model has intriguing properties: for instance, matter-antimatter asymmetry would have representation at the level of genetic code. This representation is consistent with the fact that the intron-exon boundary can occur in the middle of the codon.

2. An alternative highly attractive option is inspired by the notion of dark DNA [L1, K8]. The model for dark proton predicts under rather general assumptions that the states of dark proton correspond to those of DNA, RNA, tRNA, amino-acids. The model also predicts vertebrate genetic code correctly. This leads to quite far reaching vision about the role of dark nuclear physics in biology and evolution. Dark proton consists of three quarks (uud) but since the quarks are entangled one cannot assign a decomposition to counterparts of nucleotides to the dark proton.

   The observation of Hu and Wu [J2] related to the magnetic properties of cell membrane inspire the proposal that the dark proton sequences representing DNA sequences are associated with the lipid layers of the cell membrane as analog of double DNA strand [K18]. Unfortunately this identification is not consistent with the finding that exon-intron distinction is defined at the level of nucleotides rather than codons.

Can one save the latter option, and could both options be mutually consistent? In many-sheeted space-time nucleotides and codons naturally correspond to different space-time sheets being analogous to quark and hadron space-time sheets. Hence the two models could be consistent.

What distinguishes between the lipids to which nucleotides and codons connect? Do codons connect to sequences of three lipids such that each of them connects to nucleotide of the codon? Or do nucleotides connect to the lipids of the nuclear membrane and codons to the lipids of the cell membrane. The latter option looks rather elegant and would also provide an answer to the question "Nuclear or cell membranes or both?". Note that one can also consider memetic codons realised as sequences of 21 DNA codons connected by flux tubes to sequences of 21 dark protons as even higher level structures. This kind of flux tubes could connect DNAs of nuclei of different cells.

5.2.2 What forces exons and introns to contain integer multiple of 3 nucleotides?

What could be the physical mechanism forcing the total length of the exonic and intronic sections to be integer multiples of codons?

1. A possible mechanism is based on the existence of a phase gradient along the DNA sequence such that the increments of the phase Φ for exons of a given gene would sum up to an integer multiple of $\Delta\Phi_{tot} = n2\pi$. For intronic portions the phase gradient could vanish- The phase gradient could correspond to the phase difference associated with the weakly coupled super

conductors formed by the two DNA strands or the phases of the supra currents flowing in DNA strands.

Phase gradient is naturally accompanied by an electric field and electric fields parallel to linear molecules appear frequently in biology: DC currents of Beck indeed require the existence of electric voltage along linear structure in question [L2]. Thus flux tubes would carry both electric and magnetic fields for exon portions and could be purely magnetic for intron portions.

This mechanism allows a variant for which only exon portions are accompanied by magnetic flux tubes carrying also electric field.

2. The phase gradient must be transferred also to mRNA and splicing mechanism must detect its presence in order to cut away the portions of mRNA for which the phase gradient vanishes. This would mean that also flux tubes marking gene portions must be transcribed.

3. An attractive possibility is that the phase gradient is related with the helical structure of DNA double helix: the minimum number of DNA triplets giving rise to a multiple of 2π rotation of the helix is 10 and corresponds to the p-adic length scale $L(151)$ defining the thickness of the cell membrane. The helix can be characterized by the tangent vector of helix having axial and azimutal components:

$$K = (1, \frac{d\phi}{dz}) = (1, 10 \times 2\pi) \times \frac{1}{L(151)} \quad .$$

The phase gradient would be naturally a multiple of this vector

$$\frac{d\Phi}{dz} = nK \quad .$$

For $n = 10$ the increase of Φ would be 2π per single DNA triplet. For $n = 1$, $n = 2$, $n = 5$ and $n = 10$ the allowed lengths would come in multiples of periods 10, 5 and 2 and 1 DNA triplets respectively. Note that 5 triplets corresponds to the p-adic length scale $L(149)$ associated with the lipid layer of the cell membrane. The presence of this gradient would naturally define the splitting of the nucleotide sequence to DNA triplets and might be important in the dynamics of DNA translation and transcription.

5.2.3 Are the properties of the introns consistent with the proposed model?

The proposed interpretation of introns is consistent with the basic facts about them.

1. It has been found that the precise positions of introns in gene do not seem to affect the gene expression. Introns can start in the middle of protein building block or even in the middle of codeword.

The content of intron DNA does not correlate with the content of protein DNA. This is just what the model predicts. Introns are known to wander around genome and between cells. One could interpret their presence as some kind of experimentation with small modifications of the quantum software. The addition of intronic section to the gene does not in general have any dramatic consequences as far as gene expression is considered. If DNA acts as topological quantum computer, the addition of intronic portion would be like addition of software to a computer: a topological quantum computer program defined by the braiding assignable to the intronic section of DNA.

An interesting question is how much intron distribution affects genetic during the evolution of individual and how large the differences between members of species are.

2. In the proposed picture the evolutionary step leading from prokaryotes to eukaryotes was the emergence of the introns and gene programs having also intronic mRNA as output. This was perhaps necessitated by the emergence of the cell nucleus since introns were needed as input by the genetic program transferring mRNA out of nucleus. The emergence of the cell nucleus as $L(163)$ structure could in principle have occurred already for a cell of size $L(k) = 167$.

It seems that in cells of animals the size of the nucleus corresponds to $L(163)$ or perhaps even $L(167)$. Cell nucleus is however not all that is needed for the emergence of introns: the model requires the presence of $k = 169$ level in the hierarchy.

3. What is beautiful is that homing [I6] and retrohoming [I4] phenomena can be regarded as modification of the genetic program in a manner which is automatically internally consistent! The addition of intron does not spoil the running of the genetic program. One can quite well consider the possibility that organisms are continually experimenting with various modifications of genetic program. It might even be that homing and retrohoming phenomena make possible the evolution of the genetic program during the lifetime of individual.

If the entire mRNA corresponds to intronic portion of DNA, it remains inside nucleus. These DNA sequences might appear in cells and the mRNA in question could be seen as a remnant of RNA era. Part of silent DNA could correspond to purely intronic genes.

1. Histones form the basic protein building block of chromosomes. Histone genes are known to contain no introns. The interpretation is that the gene coding histone is in exon state permanently. If histone genes contained intron parts, histones could act as repressor genes and regulate gene activity. Since histones appear as building block proteins of nucleosomes of chromosomes, this would mean that the programs of operating system would depend on the quantum software defined by introns: obviously a highly undesired situation. Also interferons are known to have no introns. The explanation might be similar. This could explain also why the DNA of mitochondria contains no introns.

2. It is known that the pseudogenes obtained by splitting the introns are not active. For instance, the noncoding RNA and mRNA coded by gene is not translated and is left inside the nucleus. The simplest interpretation is that intronic portions define topological quantum computer programs need to carry out mRNA and transfer it from the nucleus. This interpretation would suggest that the TQC: s defined by the flux tubes connecting nucleotides to nuclear membrane are responsible for this. The TQC: s responsible for transferring molecules through the cell membrane would involves braids connecting DNA and cell membrane. The hierarchy could continue to multicellular level. In this picture the braid strands would also serve as molecular highways along which the transfer would take place.

5.3 The Phenomenon Of Superimposed Genes

Before the discovery of the structure of the genome of $\phi X174$, it was thought that the Central Dogma is absolute truth. It has however turned out that genes within genes and even overlapping genes are possible [I3]. Bacteriophage $\phi X174$, which is virus with single stranded circular DNA, contains two genes, denoted by A and D, containing genes B and E within them. A gene contains also gene A^* which starts in the middle of A and ends in the same codeword. Also the translation of a gene overlapping with A and gene C next to it have been observed in G4 bacteriophage having genome very similar to that of $\phi X174$.

To understand how the translation of mRNA can give rise to genes inside genes, one is forced to introduce the notion of reading frame shift. This concept is somewhat ad hoc since it requires that the translation of a gene within gene does not obey the usual rules since reading is stopped without stopping sign.

The geometric realization of the subprogram structure might make possible to understand gene superimposition.

1. Geometrically gene would be a linear model for its expression domain obtained by thinning it to DNA thread. Two subsequent genes G_1 and G_2 are like two subsequent body parts. If gene superposition occurs, there is third body part having overlap with both and "glued" to both: kind of joint connecting two body parts.

2. In computer analogy these "body parts" correspond would represent subprograms. Transcription generates protein which refers to some other program module realized as "body part". If the requirement that each "body part" generates program call to existing "body part", is satisfied, then gene superimposition is possible. This requirement is actually very

natural consistency condition. That our body is full of this kind of joints would suggest that gene superposition is a general phenomenon.

A concrete model is obtained in terms of many-sheeted DNA and many-sheeted versions of various RNAs. As a matter of fact, the notion of many-sheeted DNA conforms with what the picture based on the notion of magnetic body having hierarchical structure with flux quanta inside flux quanta and flux sheets traversing through sub-units of DNA sequence, suggests.

5.4 Possible Explanations Of The Silent DNA

Genome contains large amounts of silent DNA which is not transcribed. One can consider several explanations of silent DNA in TGD framework.

1. Each gene consists of a transcribed part and control part contained in region between gene and its predecessor. It might be that highly repetitive DNA located near the centromere corresponds actually to a control part of DNA and is therefore not transcribed. Enhancers and silencers could be in question and they could have contacts to larger space-time sheets and take care of the control in long length scale enhancing or silencing a large number of genes simultaneously.

2. Genome can be regarded as a large collection of program modules calling each other. Large programs typically contain a great number of modules not used by the average user. There is also a larger number of program modules whose output is not visible to the user. Silent DNA could correspond to program modules of this kind. The counterpart of unused program modules are genes which are permanently repressed. This kind of permanent repression certainly occurs during differentiation and most of DNA in given part of organism is this kind of DNA.

3. Silent DNA could also correspond to purely intronic genes which correspond to dead ends of the genetic program and are decoupled from genetic program by selection.

4. The most radical possibility is that silent DNA corresponds to genes which are expressed non-chemically and corresponding control regions affected by non-chemical transcription factors. A possible test for the existence of non-chemical gene expression is to modify the silent part of, say, neural genome and find whether and how this affects the behavior of the organism.

5. The vision about DNA as topological quantum computer [K4] had not emerged as I wrote the above list for the first time. This vision would identify silent DNA as part of software whereas genes in this framework could be seen as hardware. What makes this option especially attractive is that the detailed nucleotide composition of the DNA strand is not crucial for the functioning of the TQC program: what matters mostly is the braiding of the flux tubes connecting nucleotides/codons to lipids. The nucleotide decomposition brings only effective coloring of the braid strands which causes additional delicate effects. Therefore the braidings associated with the repetitive structures suggesting strongly the interpretation as "junk" can quite well serve as TQC programs.

5.5 About Genetic Evolution

The proposed general model of genetic program provides nontrivial insights to evolution of genome.

5.5.1 ORP and the structure of the genetic program

An interesting question is whether "ontogeny recapitulates phylogeny" principle in its original form (to be distinguished by its TGD based analog applying much more generally) could give nontrivial constraints on the structure of the genetic program.

1. One could argue that the structure of the genetic program must reflect its evolution. In the beginning of development only the lowest level genes are activated and in turn activate more evolved genes which in turn $\cdots$. The gradual emergence of new hierarchy levels would

have interpretation as emergence of new abstraction levels: statements about statements about... are formed. These levels correspond to emergence of higher level selves having more abstracted experiences about the state of organism or its organs. Geometrically the new levels would correspond to the appearance of new space-time sheets in the hierarchy of space-time sheets assignable to the magnetic body of the genome.

2. One can also defend quite different point of view. The evolution starts from simple main program corresponding to, say the length scale of a fertilized egg. Gradually subprograms corresponding to the emergence of smaller length scale structures are activated. In growth stage simple replication of the basic structure together with the activation of the lower level programs giving rise to differentiation occurs. This leads to join along boundaries/flux tube condensates of the fundamental expression domains having some finite size determined by the metabolic resources available and by self-organization. In fact, the general structure of embryogenesis supports this view whereas evolution of more complex organisms supports the first option.

3. Option 2) suggests that genetic program modules call only modules associated with shorter length scales so that higher levels cannot be activated by program call. If this is the case new levels should emerge when some space-time sheet associated with DNA expands in phase transition like manner and fuses with corresponding space-time sheets associated with neighboring cells. These phase transitions would represent the self-organization aspect of development.

4. Cell differentiation could be understood as resulting from the branching of genetic program in position dependent manner caused by diffusion gradients of transcription factors. For instance, these transcription factors would correspond to hormones which bind to receptors to form a complex binding to gene.

5.5.2 Homeostasis, loops, tautologies

If P then P is the simplest loop one can imagine and corresponds to a gene coding protein activating the gene itself. Biologically this represents endless cancer like growth limited only by lifetime of protein and resources and hence possible breakdown of the system! More complicated statement structures of this type represent n-fold tautologies: *If* P_1 *then* P_2 , *If* P_2 *then* P_3 , ..., *If* P_n *then* P_1 . These systems are also self-amplifying and correspond to a cyclic reaction in which genes code enzymes activating other genes in the cycle.

This kind of cycles might be involved with the very early evolution of life. For instance, DNA-RNA-protein trinity might have developed from a situation in which RNA coded protein which catalyzed both the reverse transcription of RNA to DNA and transcription of DNA to RNA. Thus simplest life form would have represented tautology *If P then P* ! Reverse transcription such that the reverse transcriptase also catalyzes transcription is indeed known to occur. Certain plant viruses contain inverse transcriptase synthetizing in infected cell DNA from its RNA and then DNA complementary to this. This double strand joins to the genome of the host cell and duplicates itself.

More complicated cycles involving several equivalent statements would have evolved gradually (note the parallel with the generation of mathematical theorems stating equivalence of statements!). Bacteria replicate endlessly and might perhaps be regarded as example of life form which corresponds to n-fold tautology.

5.5.3 The role of chromosomes

An interesting question is whether the organization of the genome to chromosomes could have some deeper organizational meaning. The fact that chromosome fusions and splittings are possible, suggests that chromosomes as a whole cannot have direct identification in terms of any body structure. Indeed, the realization of genetic program in terms of protein concentrations representing inputs and outputs of genes interpreted as subprograms is very flexible as far as the location of gene subprograms is considered. Only genes which form larger program structures should form geometrically connected units and the relative locations of these units could be rather free. As already found, genes seem indeed form clear geometrical subunits.

Table 3: Chromosome numbers for some species.

Animal	Man	Chimpanzee	Cow	Dog	Cat
N_c	23	24	30	39	19
Animal	Horse	Rat	Rabbit	Alligator	Frog
N_c	32	21	22	16	13
Animal	House fly	Fruit fly	Honeybee	Flatworm	Harpalinae
N_c	6	4	16	8	$18+x$

Chromosomes could be identified as a set of mutually interacting parallelly running genetic programs. One can of course consider the possibility that chromosomes mean the composition of body part to N_c linear structures, N_c being the number of chromosome pairs. Parallel interacting processing would bind these structures to single coherent hole. This division to N_c parts should be detectable at all levels of body organization. One can also consider the possibility is that the structure of chromosome parallels the structure of body and that centrosome corresponds in some sense to the brains of chromosome and the branches of chromosome correspond to right and left halves of the body. Genetic programs associated with different chromosomes are known to run in very precise synchrony. Many -sheeted DNA could explain this synchrony naturally as resulting from the interaction of genes with with classical em fields with space-time sheet containing the chromosomes.

Table 4 gives chromosome numbers for some animals. For the home fly the number of chromosome pairs is 6 whereas the number of the p-adic hierarchy levels determined by the size of the home fly is 12. For fruit fly the number of chromosome pairs is 4. Horse has 32 pairs of chromosomes and dog has 39 pairs of chromosomes whereas Homo sapiens has 23 chromosomes. The large number of chromosomes can be understand as a manner to produce large number of outcomes in breeding. The number of possible combinations of chromosomes in sexual breeding is 2^{N_c}, 2^{16} more than in case of Homo Sapiens. There are indeed very many different looking dogs barking around! The number of chromosomes varies wildly. For instance, the number of chromosomes in Harpalinae is $18+x$ [I1]! The size of this insect is about one centimeter.

The natural expectation is that the size of chromosomes has gradually grown during evolution when new p-adic space-time sheets have emerged. This process could correspond to insertion of introns to the basic DNA. The possibility coming first in mind is that the value of k_G in genes of given chromosome tends to increase as a function of the distance from the second end of gene. Genetic program realized in terms of many-sheeted space-time concept does not however require this. The genes involved in the coding the structure of given body part are known to be linearly arranged according to the structure of body part itself. This is certainly consistent with TGD picture in which body parts grow from space-time sheets associated with DNA. It seems that the highest activated space-time sheets in genes must correspond to brain, in particular frontal lobes, in case of human. This would suggest that also in chromosome the largest active length scales correspond to the region around centrosome.

5.5.4 p-Adic evolution of DNA

p-Adic evolution should involve two aspects.

1. The increase of the p-adic length scale characterizing the basic DNA modules. This suggest the classification of the basic building blocks of the genome by the p-adic length scale associated with the corresponding DNA sequences.

2. The fractal evolution involving emergence of longer p-adic length scales characterizing the size of the space-time sheets to which basic DNA sequences had # contacts. Thus the lengths of introns and exons are not expected to correlate with the p-adic scale of the space-time sheet to which they possibly have # contacts. Rather, same gene can have # contacts to arbitrarily large space-time sheets.

Consider first the critical lengths of the basic program modules. The lengths $L(149), L(151), L(157), ...$ of gene or DNA sequence might mean the emergence of something genuinely new in the evolution. This length scale hierarchy expressed in terms of $L(137)$ comes in powers of 2 as $N_{137} = 1, 2, 64, 128, 2^{10}, 2^{13}, 2^{15}, ...$

Single nucleotide pair corresponds to in double helix to distance of .34 nanometers which is larger than the length scale of $L(139)$. The structure of the double helix is such that there is a periodicity of 3.4 nanometers: this means that basic period corresponds to 10 nucleotides. This implies that 5 DNA triplets correspond to a length of 5.05 nanometers, which equals to p-adic length scale $L(149)$ if $L(151)$ is defined to be $L(151) = 10.2$ nm. $L(149)$ corresponds to the thickness of the lipid layer of cell membrane and $L(151)$ corresponds to 10 DNA triplets, to the thickness of the cell membrane and the basic period of DNA sequence when DNA triplet is regarded as a basic unit. Perhaps this periodicity is not accident but has deeper meaning possibly related to the periodicity of phase variable associated with DNA. The lengths of DNA sequences corresponding to p-adic length scale $L(k)$, $p \simeq k$, k power of prime are $N(DNA) = 2^{k-149} \times 5$ DNA triplets.

This means that the critical numbers of DNA triplets possible leading to the emergence of qualitatively new properties of organism are given by

$$\begin{aligned} N(DNA) &= 2^{(k-149)/2} \times 5 \ , \\ k &\in \{149, 151, 157, 163, 167, 169, 174, 179, 181, 191, 193, ..\} \end{aligned} \tag{5.1}$$

The few lowest critical values of DNA triplets in gene are

$$N(DNA) = n \times 5 \ ,$$
$$n = 1, 2, 2^4 = 16, 2^7 = 128, 2^9 = 512, 2^{10} = 1024, 2^{12}, 2^{15}, 2^{16}, \cdots$$

The steps of this hierarchy resembles bring in mind the evolution for the length of the basic memory unit of computer memory! One must however notice that 5 DNA triplets seems to serve as a basic unit.

The emergence of new p-adic length scales could have meant emergence of new levels of modularization in the genetic program and it is interesting to look these numbers from this point of view.

1. One could think that short sequences of precursors of DNA, mRNA and tRNA molecules were generated spontaneously by self-assemly. This implied automatically the generation of amino-acids by the more primitive counterparts of transcription and translation processes. The lengths of DNA molecules began gradually grow and at the critical lengths of DNA corresponding to p-adic length scales dramatic new effects emerged. Also new space-time sheets emerged in the genome and the first guess is that this occurred for the critical sizes of the organism given by p-adic length scales.

2. Formation of lipid layers might have been the revolution occurring at this stage and since lipids should have had size of order $L(149)$. This revolution should have occurred when the length of the genome became longer than 5 DNA triplets and meant formation of lipid layers by self organization process known to occur in all liquid crystals: these layers were perhaps formed in the surface of water such that hydrophobic ends of proteins would have pointed out of water. Self organization presumably led simultaneously to the formation of double membranes having thickness $L(151)$ such that the hydrophobic ends of proteins pointed in the interior of the double membrane. Second revolution became possible when the number of DNA triplets became larger than 10 triplets so that proteins connecting cell interior of the double membrane to its exterior became possible and the control of ion concentrations became in principle possible. Transfer RNA (tRNA) has length of at most 27 triplets. Third revolution should have occurred $L(157)$, which corresponds to 80 triplets.

3. Smallest viruses possessing single strand of DNA have lengths between 15-100 nanometers and this suggest that genome correspond to p-adic length scales $L(149)$, $L(151)$ and $L(157)$. These length scales could characterize largest space-time sheets also present in genome. The building blocks of the envelope of viruses are genetically coded separately and self-assemble

spontaneously so that only building blocks need to be coded. Therefore p-adic prime associated with the genome of virus could be smaller than that determined by the size of the virus. Viruses with two DNA strands have sizes between $250 - 1000$ nanometers. This suggest that the emergence of $k = 163$ length scale in the genome of virus was accompanied by the emergence of double stranded DNA. $k = 163$ is perhaps the largest p-adic length scale associated with virus genome.

4. Bacteria have typically sizes of $1 - 10$ micro-meters. This suggests that $k = 163, 167, 169$ are the possible space-time sheets associated with the bacterial genome. The emergence of $k = 169$ could have meant the emergence of multicellulars and generation of epithelial sheet like structure consisting of two cell layers as well as emergence of introns and DNA cognition.

Consider now the typical lengths for the structures of the eukaryotic genome.

1. The presence of introns means that the length of a gene coding given protein plus introns is much longer than the DNA coding only the protein. The higher the evolutionary level of species, the larger the fraction of introns. For human genome the fraction of exons is roughly 1 per cent. The typical length of hnRNA in nucleus is 6.000-8.000 np (nucleotide pairs) which corresponds to 18 micro-meters and length scale $L(163)$ and $L(167)$. Even genes with length 20.000 np are possible and correspond to $L(169)$. The lengths of mRNA vary between 500-3.000 nucleotides corresponding to interval 1.7×10^{-7}-10^{-6} meters and length scales $L(157)$ and $L(163)$. RNA sequences coding typical protein consisting of roughly 300 amino acids are about 3×10^{-7} meters and correspond to $L(159)$.

2. Most of the highly repetitive DNA has rather short length between $5 - 300$ nucleotides. Introns having typically lengths between $10 - 1000$ nucleotide pairs. The length of ribosomal DNA is not longer than 10^3 nucleotides. These examples suggests that the basic program modules correspond to p-adic length scales between $L(139)$ and $L(157)$ and that introns and genes are built as fractal versions of the basic program modules possibly present in all plants and animals. The basic programs are chemically identical. They could however have wormhole contacts to increasingly larger space-time sheets so that organism possesses fractal like structural hierarchy. Alternatively, the contacts are on space-time sheets with same p in all animals but the sizes of the join along boundaries/flux tube condensates formed by fundamental expression domains depend on organism. The frequent occurrence of Hox genes in the genetic code of body parts of various sizes in the entire animal kingdom is consistent with both options.

6 Ideas About Concrete Realization Of Genetic Programs

In this section some ideas about concrete realization of genetic programs are discussed. The ideas derive from the first years of TGD inspired quantum biology and it would be interesting to see how the systematic application of recent understanding of TGD could enrichen the picture.

6.1 How Gene Expression Is Regulated?

In case of prokaryotes the regulation of transcription is quite satisfactorily understood. The problem is to understand how transcription is regulated in case of the eukaryotes and here the notion of many-sheeted DNA could be crucial.

6.1.1 Operon theory for the regulation of gene expression in prokaryotes

Jacob, Monod and Pardee [I3] suggested operon theory for the regulation of the transcription of genes responsible for lactose production in E. coli. The presence of lactose induces E. coli to produce 3 enzymes needed in the production of lactose. The enzymes correspond to three structure genes x, y, z of lactose. The mechanism is following.

1. So called i gene regulates the production of the building block proteins of the repressor protein which self-assembles as tetramer of the repressor protein.

2. Repressor protein binds to a specific site next to the promotor and hinders the binding of the RNA polymerase so that transcription becomes impossible.

3. Inductor, in the present case lactose, binds to the repressor protein and hence hinders the formation of the repressor tetrameres so that transcription becomes possible.

4. The transcript contains not only the gene but also the entire operon so that several genes are translated simultaneously.

6.1.2 How eukaryotes differ from prokaryotes?

Operon theory does not generalize as such to eukaryotes. Although the notion of the promoter generalizes, there is no clear-cut evidence for operons [I3]. Rather, silencers and enhancers could take the role of the inducers and repressors in the eukaryotic gene expression. The action of a silencer/enhancer is not sensitive to its precise location or orientation and the distance from promoter can be more than thousand nucleotide pairs. TGD based explanation is based on the notions of many-sheeted DNA and protein. Silencers and enhancers mediate the interaction of the atomic space-time sheet of DNA with the classical fields of some larger space-time sheet. This interaction makes possible top-bottom type control analogous to the control of slave by master in Haken's theory of self-organization. Both classical em and Z^0 fields can control the gene expression in this manner. In eukaryotes classical Z^0 fields could have especially important role. Classical Z^0 fields are believed to be crucial for the model of cognition and this suggests that "mind-matter" interaction could at least partially relay on classical Z^0 fields and enhancers/silencers. Silencers and enhancers could make also possible Josephson junctions between gene space-time sheet and some larger space-time sheet and thus realize "biofeedback".

Second difference is related to introns. It is known that introns and exons can change their roles and it is known that there are several pathways for splicing leading to different proteins, isoforms. The replacement of single-valued gene$\rightarrow$ polypeptide map with many-valued map obviously increases the information content of gene. The interpretation of introns and exons as two interspersed computer codes with intron lines of code separated from exons by "comment signs" marking each nucleotide of intron is attractive model for the situation. Dropping of some comment signs changes the result of the splicing process. Comment sign distribution could be dynamical and tissue specific so that one could say that genome is not invariant of species but only of a particular tissue type. This obviously reduces the genetic determinism and gives organism better abilities to survive. In human genome 1 percent of gene corresponds to exons in the "normal" state (in whatever manner that state is defined). The number of various combinations of exon and intron combinations is 2^N where N denotes the numbers of basic components of gene (perhaps coding proteins having no decomposition to modular proteins). The number of combinations increases exponentially with N and provides huge flexibility.

The interpretation of various exon-intron decompositions as statements of Boolean algebra of statements about N basic statements suggests strongly itself. In particular, exonic and intronic proteins for same intron-exon decomposition would naturally correspond to a statement and its negation. Thus one could regard eukaryotes as representing higher levels in the hierarchy of abstractions in which prokaryotes represent the lowest level. Eukaryotic genome would be a collection of genes identifiable as Boolean algebras and the running of the genetic program would mean that at given moment some statement in some of these Boolean algebras is experienced to be true. This kind of identification would mean effectively understanding of the logical meaning of the genetic code. The dominance of particular exon-intron decomposition over its complement would simply mean that this Boolean statement is true while its complement is not true. This suggests the possibility that only those 2^{N-1} exon-intron configuration which represent statements consistent with a given atomic statement are usually realized (atomic statement corresponds to a one-element subset in the set theoretic representation of Boolean algebra).

The fact that introns and exons are set theoretical complements of each other raises the possibility that the proteins coded by introns and exons have opposite effects as transcription factors. For instance, exons could code enhancer and introns could code silencer. This implies that same gene can act as both enhancer and silencer. The only thing needed is that the roles of introns and exons are changed. When this occurs, the production of intron-RNA complexes begins and the production of the composite protein coded by introns acting as exons stops. When the production

of silencer protein coded by introns ceases, the silencer proteins associated with the operator sites gradually decay and gene expression also enhanced by enhancer proteins can start. Thus the activation of the gene module activates automatically the gene modules which it calls and genetic program runs.

This would suggests rather general mechanism of gene expression.

1. There exists 2^{N-1} exon-intron decomposition plus their complements. Depending on the state of gene with given exon-intron decomposition, exons or introns are translated to protein. Both introns and exons represent statements. The modular decomposition of protein to subproteins often represented by the decomposition to exons corresponds to the decomposition of the statement E to $E = E_1 \& E_2 ... \& E_n$. Same holds true for I: $I = I_1 \& I_2 ... \& I_n$.

2. The proteins coded usually by exons activate some genes and these proteins appear as prerequisites of type

$$IF\ [E_1\ OR\ E_2 \cdots\ OR\ E_n]\ THEN \cdots\ .$$

   The presence of all activators is not necessary.

3. Introns correspond to repressors quite generally. The proteins coded by introns in the "abnormal" state of the gene correspond to prerequisites of type

$$\cdots IF\,[NOT(I_1) \& NOT(I_2) ... \& NOT(I_n)]\ THEN \cdots\ .$$

   This means that the absence of all repressors from operator site is necessary. The general form of gene statement is

$$IF\ [E_1\ OR\ E_2 \cdots\ OR\ E_n] \& IF\,[NOT(I_1) \& NOT(I_2) ... \& NOT(I_n)]\ THEN \cdots\ .$$

   Various E_k: s and I_k: s represent proteins in turn having modular decomposition to a product of more primitive statements.

6.1.3 The role of the hierarchy of Josephson currents

The control– and coordination hierarchy formed by super conductors represented by space-time sheets coupled to each other by join along boundaries bonds suggests new quantum level control mechanisms for genetic expression. Josephson currents at resonant frequencies corresponding to some magnetic transition of gene or its substructure could "wake-up" the gene self and initiate the self-organization process leading to the gene activity. Silencers and enhancers could correspond to proteins which have join along boundaries bonds/flux tube contacts to larger space-time sheets serving as masters. This would explain why neither the exact position nor orientation of the silencer or enhancer is not important for their functioning.

Various genes and and associated control structures have mutual Josephson junctions controlling gene expression. This would mean the presence of extremely weak longitudinal electromagnetic fields (the potential differences over Josephson junctions would be in 10^{-14} eV range). The control mechanisms behind morphogenesis are poorly understood and phase gradients along chromosomes and along the growing organism could be involved with the control of morphogenesis. The fact that the rate of the translation process is about 20 amino-acids per seconds is in accordance with the idea that this process is controlled by a Josephson current associated with some ion having this frequency.

6.2 Model For The Physical Distinction Between Exons And Introns

Introns seem to begin with nucleotide pair CT and end with pair TC: the assumption that this is the sole signature of the beginning of intron is however not consistent with the observation about the change of the roles of intron and exons. Furthermore, this criterion is obviously not sufficient for telling where intron begins and ends since also ordinary genes can contain similar section.

There are several constraints on the marking telling whether a given nucleotide is exon or intron.

1. The splicing mechanism or DNA telling which portions of it are transcribed to mRNA should rely on the marking mechanism. This requires that the enzymes responsible for the splicing are able to distinguish between introns and exons.

2. Introns can start and end in the middle of DNA triplet. This suggests that the marking is assigned with each nucleotide separately rather than each codon.

3. Both intronic and exon portions of DNA are transcribed to mRNA and splicing of the intronic portions of mRNA occurs only after that. Therefore the total length of the exon portions of DNA corresponding to a given gene must correspond to integer multiple of codons and there must exist some mechanism forcing this. The length of the intronic portions seems to be free.

4. The transformations of exons and introns to each other are known to occur. Tere must exist a physical mechanism changing the marking.

6.2.1 Could magnetic flux tubes serving as braid strands distinguish between introns and exons?

A priori one can imagine an endless number of possibilities for the physical realization of the binary digit distinguishing between exon and intron nucleotide. The vision about DNA as topological quantum computer with flux tubes connecting introns to the lipids of nuclear and cell membranes (not necessarily those of same nucleus or cell) acting as braid strands would realize this distinction naturally: the introduction of the connecting flux tube would transform exon an intron. Reconnection of pairs of flux tubes between DNA nucleotides or codons and lipids would allow to effectively eliminate the braid strands and transform pair of introns to exons. The first guess would be that the presence of braid strand could make it impossible to perform transcription. This is not the case: both intronic and exon portions are transcribed to mRNA and slicing process cuts off the intronic portions of mRNA after that.

1. The original model of DNA as topological quantum computer [K4] assigns to the nucleotides A, T, C, G quarks u and d and antiquarks $\overline{u}$ and $\overline{d}$. This model has intriguing properties: for instance, matter-antimatter asymmetry would have representation at the level of genetic code. This representation is consistent with the fact that the intron-exon boundary can occur in the middle of the codon.

2. An alternative highly attractive option is inspired by the notion of dark DNA [L1, K8]. The model for dark proton predicts under rather general assumptions that the states of dark proton correspond to those of DNA, RNA, tRNA, amino-acids. The model also predicts vertebrate genetic code correctly. This leads to quite far reaching vision about the role of dark nuclear physics in biology and evolution. Dark proton consists of three quarks (uud) but since the quarks are entangled one cannot assign a decomposition to counterparts of nucleotides to the dark proton.

   The observation of Hu and Wu [J2] related to the magnetic properties of cell membrane inspire the proposal that the dark proton sequences representing DNA sequences are associated with the lipid layers of the cell membrane as analog of double DNA strand [K18]. Unfortunately this identification is not consistent with the finding that exon-intron distinction is defined at the level of nucleotides rather than codons.

Can one save the latter option, and could both options be mutually consistent? In many-sheeted space-time nucleotides and codons naturally correspond to different space-time sheets being analogous to quark and hadron space-time sheets. Hence the two models could be consistent.

What distinguishes between the lipids to which nucleotides and codons connect? Do codons connect to sequences of three lipids such that each of them connects to nucleotide of the codon? Or do nucleotides connect to the lipids of the nuclear membrane and codons to the lipids of the cell membrane. The latter option looks rather elegant and would also provide an answer to the

question "Nuclear or cell membranes or both?". Note that one can also consider memetic codons realised as sequences of 21 DNA codons connected by flux tubes to sequences of 21 dark protons as even higher level structures. This kind of flux tubes could connect DNAs of nuclei of different cells.

6.2.2 What forces exons and introns to contain integer multiple of 3 nucleotides?

What could be the physical mechanism forcing the total length of the exonic and intronic sections to be integer multiples of codons?

1. A possible mechanism is based on the existence of a phase gradient along the DNA sequence such that the increments of the phase Φ for exons of a given gene would sum up to an integer multiple of $\Delta\Phi_{tot} = n2\pi$. For intronic portions the phase gradient could vanish- The phase gradient could correspond to the phase difference associated with the weakly coupled super conductors formed by the two DNA strands or the phases of the supra currents flowing in DNA strands.

   Phase gradient is naturally accompanied by an electric field and electric fields parallel to linear molecules appear frequently in biology: DC currents of Beck indeed require the existence of electric voltage along linear structure in question [L2]. Thus flux tubes would carry both electric and magnetic fields for exon portions and could be purely magnetic for intron portions.

   This mechanism allows a variant for which only exon portions are accompanied by magnetic flux tubes carrying also electric field.

2. The phase gradient must be transferred also to mRNA and splicing mechanism must detect its presence in order to cut away the portions of mRNA for which the phase gradient vanishes. This would mean that also flux tubes marking gene portions must be transcribed.

3. An attractive possibility is that the phase gradient is related with the helical structure of DNA double helix: the minimum number of DNA triplets giving rise to a multiple of 2π rotation of the helix is 10 and corresponds to the p-adic length scale $L(151)$ defining the thickness of the cell membrane. The helix can be characterized by the tangent vector of helix having axial and azimutal components:

   $$K = (1, \frac{d\phi}{dz}) = (1, 10 \times 2\pi) \times \frac{1}{L(151)} \quad .$$

   The phase gradient would be naturally a multiple of this vector

   $$\frac{d\Phi}{dz} = nK \quad .$$

   For $n = 10$ the increase of Φ would be 2π per single DNA triplet. For $n = 1$, $n = 2$, $n = 5$ and $n = 10$ the allowed lengths would come in multiples of periods 10, 5 and 2 and 1 DNA triplets respectively. Note that 5 triplets corresponds to the p-adic length scale $L(149)$ associated with the lipid layer of the cell membrane. The presence of this gradient would naturally define the splitting of the nucleotide sequence to DNA triplets and might be important in the dynamics of DNA translation and transcription.

6.2.3 Are the properties of the introns consistent with the proposed model?

The proposed interpretation of introns is consistent with the basic facts about them.

1. It has been found that the precise positions of introns in gene do not seem to affect the gene expression. Introns can start in the middle of protein building block or even in the middle of codeword.

   The content of intron DNA does not correlate with the content of protein DNA. This is just what the model predicts. Introns are known to wander around genome and between cells. One could interpret their presence as some kind of experimentation with small modifications

of the quantum software. The addition of intronic section to the gene does not in general have any dramatic consequences as far as gene expression is considered. If DNA acts as topological quantum computer, the addition of intronic portion would be like addition of software to a computer: a topological quantum computer program defined by the braiding assignable to the intronic section of DNA.

An interesting question is how much intron distribution affects genetic during the evolution of individual and how large the differences between members of species are.

2. In the proposed picture the evolutionary step leading from prokaryotes to eukaryotes was the emergence of the introns and gene programs having also intronic mRNA as output. This was perhaps necessitated by the emergence of the cell nucleus since introns were needed as input by the genetic program transferring mRNA out of nucleus. The emergence of the cell nucleus as $L(163)$ structure could in principle have occurred already for a cell of size $L(k) = 167$. It seems that in cells of animals the size of the nucleus corresponds to $L(163)$ or perhaps even $L(167)$. Cell nucleus is however not all that is needed for the emergence of introns: the model requires the presence of $k = 169$ level in the hierarchy.

3. What is beautiful is that homing [I6] and retrohoming [I4] phenomena can be regarded as modification of the genetic program in a manner which is automatically internally consistent! The addition of intron does not spoil the running of the genetic program. One can quite well consider the possibility that organisms are continually experimenting with various modifications of genetic program. It might even be that homing and retrohoming phenomena make possible the evolution of the genetic program during the lifetime of individual.

If the entire mRNA corresponds to intronic portion of DNA, it remains inside nucleus. These DNA sequences might appear in cells and the mRNA in question could be seen as a remnant of RNA era. Part of silent DNA could correspond to purely intronic genes.

1. Histones form the basic protein building block of chromosomes. Histone genes are known to contain no introns. The interpretation is that the gene coding histone is in exon state permanently. If histone genes contained intron parts, histones could act as repressor genes and regulate gene activity. Since histones appear as building block proteins of nucleosomes of chromosomes, this would mean that the programs of operating system would depend on the quantum software defined by introns: obviously a highly undesired situation. Also interferons are known to have no introns. The explanation might be similar. This could explain also why the DNA of mitochondria contains no introns.

2. It is known that the pseudogenes obtained by splitting the introns are not active. For instance, the noncoding RNA and mRNA coded by gene is not translated and is left inside the nucleus. The simplest interpretation is that intronic portions define topological quantum computer programs need to carry out mRNA and transfer it from the nucleus. This interpretation would suggest that the TQC: s defined by the flux tubes connecting nucleotides to nuclear membrane are responsible for this. The TQC: s responsible for transferring molecules through the cell membrane would involves braids connecting DNA and cell membrane. The hierarchy could continue to multicellular level. In this picture the braid strands would also serve as molecular highways along which the transfer would take place.

6.3 The Phenomenon Of Superimposed Genes

Before the discovery of the structure of the genome of $\phi X174$, it was thought that the Central Dogma is absolute truth. It has however turned out that genes within genes and even overlapping genes are possible [I3]. Bacteriophage $\phi X174$, which is virus with single stranded circular DNA, contains two genes, denoted by A and D, containing genes B and E within them. A gene contains also gene A^* which starts in the middle of A and ends in the same codeword. Also the translation of a gene overlapping with A and gene C next to it have been observed in G4 bacteriophage having genome very similar to that of $\phi X174$.

To understand how the translation of mRNA can give rise to genes inside genes, one is forced to introduce the notion of reading frame shift. This concept is somewhat ad hoc since it requires

that the translation of a gene within gene does not obey the usual rules since reading is stopped without stopping sign.

The geometric realization of the subprogram structure might make possible to understand gene superimposition.

1. Geometrically gene would be a linear model for its expression domain obtained by thinning it to DNA thread. Two subsequent genes G_1 and G_2 are like two subsequent body parts. If gene superposition occurs, there is third body part having overlap with both and "glued" to both: kind of joint connecting two body parts.

2. In computer analogy these "body parts" correspond would represent subprograms. Transcription generates protein which refers to some other program module realized as "body part". If the requirement that each "body part" generates program call to existing "body part", is satisfied, then gene superimposition is possible. This requirement is actually very natural consistency condition. That our body is full of this kind of joints would suggest that gene superposition is a general phenomenon.

A concrete model is obtained in terms of many-sheeted DNA and many-sheeted versions of various RNAs. As a matter of fact, the notion of many-sheeted DNA conforms with what the picture based on the notion of magnetic body having hierarchical structure with flux quanta inside flux quanta and flux sheets traversing through sub-units of DNA sequence, suggests.

6.4 Possible Explanations Of The Silent DNA

Genome contains large amounts of silent DNA which is not transcribed. One can consider several explanations of silent DNA in TGD framework.

1. Each gene consists of a transcribed part and control part contained in region between gene and its predecessor. It might be that highly repetitive DNA located near the centromere corresponds actually to a control part of DNA and is therefore not transcribed. Enhancers and silencers could be in question and they could have contacts to larger space-time sheets and take care of the control in long length scale enhancing or silencing a large number of genes simultaneously.

2. Genome can be regarded as a large collection of program modules calling each other. Large programs typically contain a great number of modules not used by the average user. There is also a larger number of program modules whose output is not visible to the user. Silent DNA could correspond to program modules of this kind. The counterpart of unused program modules are genes which are permanently repressed. This kind of permanent repression certainly occurs during differentiation and most of DNA in given part of organism is this kind of DNA.

3. Silent DNA could also correspond to purely intronic genes which correspond to dead ends of the genetic program and are decoupled from genetic program by selection.

4. The most radical possibility is that silent DNA corresponds to genes which are expressed non-chemically and corresponding control regions affected by non-chemical transcription factors. A possible test for the existence of non-chemical gene expression is to modify the silent part of, say, neural genome and find whether and how this affects the behavior of the organism.

5. The vision about DNA as topological quantum computer [K4] had not emerged as I wrote the above list for the first time. This vision would identify silent DNA as part of software whereas genes in this framework could be seen as hardware. What makes this option especially attractive is that the detailed nucleotide composition of the DNA strand is not crucial for the functioning of the TQC program: what matters mostly is the braiding of the flux tubes connecting nucleotides/codons to lipids. The nucleotide decomposition brings only effective coloring of the braid strands which causes additional delicate effects. Therefore the braidings associated with the repetitive structures suggesting strongly the interpretation as "junk" can quite well serve as TQC programs.

6.5 About Genetic Evolution

The proposed general model of genetic program provides nontrivial insights to evolution of genome.

6.5.1 ORP and the structure of the genetic program

An interesting question is whether "ontogeny recapitulates phylogeny" principle in its original form (to be distinguished by its TGD based analog applying much more generally) could give nontrivial constraints on the structure of the genetic program.

1. One could argue that the structure of the genetic program must reflect its evolution. In the beginning of development only the lowest level genes are activated and in turn activate more evolved genes which in turn $\cdots$. The gradual emergence of new hierarchy levels would have interpretation as emergence of new abstraction levels: statements about statements about... are formed. These levels correspond to emergence of higher level selves having more abstracted experiences about the state of organism or its organs. Geometrically the new levels would correspond to the appearance of new space-time sheets in the hierarchy of space-time sheets assignable to the magnetic body of the genome.

2. One can also defend quite different point of view. The evolution starts from simple main program corresponding to, say the length scale of a fertilized egg. Gradually subprograms corresponding to the emergence of smaller length scale structures are activated. In growth stage simple replication of the basic structure together with the activation of the lower level programs giving rise to differentiation occurs. This leads to join along boundaries/flux tube condensates of the fundamental expression domains having some finite size determined by the metabolic resources available and by self-organization. In fact, the general structure of embryogenesis supports this view whereas evolution of more complex organisms supports the first option.

3. Option 2) suggests that genetic program modules call only modules associated with shorter length scales so that higher levels cannot be activated by program call. If this is the case new levels should emerge when some space-time sheet associated with DNA expands in phase transition like manner and fuses with corresponding space-time sheets associated with neighboring cells. These phase transitions would represent the self-organization aspect of development.

4. Cell differentiation could be understood as resulting from the branching of genetic program in position dependent manner caused by diffusion gradients of transcription factors. For instance, these transcription factors would correspond to hormones which bind to receptors to form a complex binding to gene.

6.5.2 Homeostasis, loops, tautologies

If P then P is the simplest loop one can imagine and corresponds to a gene coding protein activating the gene itself. Biologically this represents endless cancer like growth limited only by lifetime of protein and resources and hence possible breakdown of the system! More complicated statement structures of this type represent n-fold tautologies: *If* P_1 *then* P_2 , *If* P_2 *then* P_3 , ..., *If* P_n *then* P_1 . These systems are also self-amplifying and correspond to a cyclic reaction in which genes code enzymes activating other genes in the cycle.

This kind of cycles might be involved with the very early evolution of life. For instance, DNA-RNA-protein trinity might have developed from a situation in which RNA coded protein which catalyzed both the reverse transcription of RNA to DNA and transcription of DNA to RNA. Thus simplest life form would have represented tautology *If P then P* ! Reverse transcription such that the reverse transcriptase also catalyzes transcription is indeed known to occur. Certain plant viruses contain inverse transcriptase synthetizing in infected cell DNA from its RNA and then DNA complementary to this. This double strand joins to the genome of the host cell and duplicates itself.

Table 4: Chromosome numbers for some species.

Animal	Man	Chimpanzee	Cow	Dog	Cat
N_c	23	24	30	39	19
Animal	Horse	Rat	Rabbit	Alligator	Frog
N_c	32	21	22	16	13
Animal	House fly	Fruit fly	Honeybee	Flatworm	Harpalinae
N_c	6	4	16	8	$18+x$

More complicated cycles involving several equivalent statements would have evolved gradually (note the parallel with the generation of mathematical theorems stating equivalence of statements!). Bacteria replicate endlessly and might perhaps be regarded as example of life form which corresponds to n-fold tautology.

6.5.3 The role of chromosomes

An interesting question is whether the organization of the genome to chromosomes could have some deeper organizational meaning. The fact that chromosome fusions and splittings are possible, suggests that chromosomes as a whole cannot have direct identification in terms of any body structure. Indeed, the realization of genetic program in terms of protein concentrations representing inputs and outputs of genes interpreted as subprograms is very flexible as far as the location of gene subprograms is considered. Only genes which form larger program structures should form geometrically connected units and the relative locations of these units could be rather free. As already found, genes seem indeed form clear geometrical subunits.

Chromosomes could be identified as a set of mutually interacting parallelly running genetic programs. One can of course consider the possibility that chromosomes mean the composition of body part to N_c linear structures, N_c being the number of chromosome pairs. Parallel interacting processing would bind these structures to single coherent hole. This division to N_c parts should be detectable at all levels of body organization. One can also consider the possibility is that the structure of chromosome parallels the structure of body and that centrosome corresponds in some sense to the brains of chromosome and the branches of chromosome correspond to right and left halves of the body. Genetic programs associated with different chromosomes are known to run in very precise synchrony. Many -sheeted DNA could explain this synchrony naturally as resulting from the interaction of genes with with classical em fields with space-time sheet containing the chromosomes.

Table 4 gives chromosome numbers for some animals. For the home fly the number of chromosome pairs is 6 whereas the number of the p-adic hierarchy levels determined by the size of the home fly is 12. For fruit fly the number of chromosome pairs is 4. Horse has 32 pairs of chromosomes and dog has 39 pairs of chromosomes whereas Homo sapiens has 23 chromosomes. The large number of chromosomes can be understand as a manner to produce large number of outcomes in breeding. The number of possible combinations of chromosomes in sexual breeding is 2^{N_c}, 2^{16} more than in case of Homo Sapiens. There are indeed very many different looking dogs barking around! The number of chromosomes varies wildly. For instance, the number of chromosomes in Harpalinae is $18+x$ [I1]! The size of this insect is about one centimeter.

The natural expectation is that the size of chromosomes has gradually grown during evolution when new p-adic space-time sheets have emerged. This process could correspond to insertion of introns to the basic DNA. The possibility coming first in mind is that the value of k_G in genes of given chromosome tends to increase as a function of the distance from the second end of gene. Genetic program realized in terms of many-sheeted space-time concept does not however require this. The genes involved in the coding the structure of given body part are known to be linearly arranged according to the structure of body part itself. This is certainly consistent with TGD picture in which body parts grow from space-time sheets associated with DNA. It seems that the highest activated space-time sheets in genes must correspond to brain, in particular frontal lobes,

in case of human. This would suggest that also in chromosome the largest active length scales correspond to the region around centrosome.

6.5.4 p-Adic evolution of DNA

p-Adic evolution should involve two aspects.

1. The increase of the p-adic length scale characterizing the basic DNA modules. This suggest the classification of the basic building blocks of the genome by the p-adic length scale associated with the corresponding DNA sequences.

2. The fractal evolution involving emergence of longer p-adic length scales characterizing the size of the space-time sheets to which basic DNA sequences had # contacts. Thus the lengths of introns and exons are not expected to correlate with the p-adic scale of the space-time sheet to which they possibly have # contacts. Rather, same gene can have # contacts to arbitrarily large space-time sheets.

Consider first the critical lengths of the basic program modules. The lengths $L(149), L(151), L(157), ...$ of gene or DNA sequence might mean the emergence of something genuinely new in the evolution. This length scale hierarchy expressed in terms of $L(137)$ comes in powers of 2 as $N_{137} = 1, 2, 64, 128, 2^{10}, 2^{13}, 2^{15}, ...$

Single nucleotide pair corresponds to in double helix to distance of .34 nanometers which is larger than the length scale of $L(139)$. The structure of the double helix is such that there is a periodicity of 3.4 nanometers: this means that basic period corresponds to 10 nucleotides. This implies that 5 DNA triplets correspond to a length of 5.05 nanometers, which equals to p-adic length scale $L(149)$ if $L(151)$ is defined to be $L(151) = 10.2$ nm. $L(149)$ corresponds to the thickness of the lipid layer of cell membrane and $L(151)$ corresponds to 10 DNA triplets, to the thickness of the cell membrane and the basic period of DNA sequence when DNA triplet is regarded as a basic unit. Perhaps this periodicity is not accident but has deeper meaning possibly related to the periodicity of phase variable associated with DNA. The lengths of DNA sequences corresponding to p-adic length scale $L(k)$, $p \simeq k$, k power of prime are $N(DNA) = 2^{k-149} \times 5$ DNA triplets.

This means that the critical numbers of DNA triplets possible leading to the emergence of qualitatively new properties of organism are given by

$$\begin{aligned} N(DNA) &= 2^{(k-149)/2} \times 5 \ , \\ k &\in \{149, 151, 157, 163, 167, 169, 174, 179, 181, 191, 193, ..\} \end{aligned} \tag{6.1}$$

The few lowest critical values of DNA triplets in gene are

$$N(DNA) = n \times 5 \ ,$$
$$n = 1, 2, 2^4 = 16, 2^7 = 128, 2^9 = 512, 2^{10} = 1024, 2^{12}, 2^{15}, 2^{16}, \cdots$$

The steps of this hierarchy resembles bring in mind the evolution for the length of the basic memory unit of computer memory! One must however notice that 5 DNA triplets seems to serve as a basic unit.

The emergence of new p-adic length scales could have meant emergence of new levels of modularization in the genetic program and it is interesting to look these numbers from this point of view.

1. One could think that short sequences of precursors of DNA, mRNA and tRNA molecules were generated spontaneously by self-assemly. This implied automatically the generation of amino-acids by the more primitive counterparts of transcription and translation processes. The lengths of DNA molecules began gradually grow and at the critical lengths of DNA corresponding to p-adic length scales dramatic new effects emerged. Also new space-time sheets emerged in the genome and the first guess is that this occurred for the critical sizes of the organism given by p-adic length scales.

2. Formation of lipid layers might have been the revolution occurring at this stage and since lipids should have had size of order $L(149)$. This revolution should have occurred when the length of the genome became longer than 5 DNA triplets and meant formation of lipid layers by self organization process known to occur in all liquid crystals: these layers were perhaps formed in the surface of water such that hydrophobic ends of proteins would have pointed out of water. Self organization presumably led simultaneously to the formation of double membranes having thickness $L(151)$ such that the hydrophobic ends of proteins pointed in the interior of the double membrane. Second revolution became possible when the number of DNA triplets became larger than 10 triplets so that proteins connecting cell interior of the double membrane to its exterior became possible and the control of ion concentrations became in principle possible. Transfer RNA (tRNA) has length of at most 27 triplets. Third revolution should have occurred $L(157)$, which corresponds to 80 triplets.

3. Smallest viruses possessing single strand of DNA have lengths between 15-100 nanometers and this suggest that genome correspond to p-adic length scales $L(149)$, $L(151)$ and $L(157)$. These length scales could characterize largest space-time sheets also present in genome. The building blocks of the envelope of viruses are genetically coded separately and self-assemble spontaneously so that only building blocks need to be coded. Therefore p-adic prime associated with the genome of virus could be smaller than that determined by the size of the virus. Viruses with two DNA strands have sizes between $250 - 1000$ nanometers. This suggest that the emergence of $k = 163$ length scale in the genome of virus was accompanied by the emergence of double stranded DNA. $k = 163$ is perhaps the largest p-adic length scale associated with virus genome.

4. Bacteria have typically sizes of $1 - 10$ micro-meters. This suggests that $k = 163, 167, 169$ are the possible space-time sheets associated with the bacterial genome. The emergence of $k = 169$ could have meant the emergence of multicellulars and generation of epithelial sheet like structure consisting of two cell layers as well as emergence of introns and DNA cognition.

Consider now the typical lengths for the structures of the eukaryotic genome.

1. The presence of introns means that the length of a gene coding given protein plus introns is much longer than the DNA coding only the protein. The higher the evolutionary level of species, the larger the fraction of introns. For human genome the fraction of exons is roughly 1 per cent. The typical length of hnRNA in nucleus is 6.000-8.000 np (nucleotide pairs) which corresponds to 18 micro-meters and length scale $L(163)$ and $L(167)$. Even genes with length 20.000 np are possible and correspond to $L(169)$. The lengths of mRNA vary between 500-3.000 nucleotides corresponding to interval 1.7×10^{-7}-10^{-6} meters and length scales $L(157)$ and $L(163)$. RNA sequences coding typical protein consisting of roughly 300 amino acids are about 3×10^{-7} meters and correspond to $L(159)$.

2. Most of the highly repetitive DNA has rather short length between $5 - 300$ nucleotides. Introns having typically lengths between $10 - 1000$ nucleotide pairs. The length of ribosomal DNA is not longer than 10^3 nucleotides. These examples suggests that the basic program modules correspond to p-adic length scales between $L(139)$ and $L(157)$ and that introns and genes are built as fractal versions of the basic program modules possibly present in all plants and animals. The basic programs are chemically identical. They could however have wormhole contacts to increasingly larger space-time sheets so that organism possesses fractal like structural hierarchy. Alternatively, the contacts are on space-time sheets with same p in all animals but the sizes of the join along boundaries/flux tube condensates formed by fundamental expression domains depend on organism. The frequent occurrence of Hox genes in the genetic code of body parts of various sizes in the entire animal kingdom is consistent with both options.

7 TGD Inspired Ideas About The Regulation Of Morphogenesis

The understanding of morphogenesis provides a challenge for the TGD inspired notion of the many-sheeted DNA. The difficult task is to separate chemistry from geometry and identify those features of morphogenesis necessitating the concept of the many-sheeted DNA. Also the role of quantum control mechanisms must be understood. In the sequel general ideas about the quantum control of morphogenesis are discussed and a very brief review about the morphogenesis in Drosophila is given to provide bird's eye of view about genetic control. Also Hox genes and TGD based model for Hox genes is discussed.

7.1 Biological Alarm Clocks And Morphogenesis

Gene expression is known to involve non-chemical transcription factors (enhancers silencers), whose underlying interaction mechanism is not not known. TGD suggest an extremely general mechanism of quantum control and coordination based on Josephson currents flowing between space-time sheets representing various levels of the biological self-hierarchy [K12, K13]. Josephson currents themselves act as clocks. In case that the potential difference over Josephson junction corresponds to the difference of the energies for the states of charge carriers localized inside either super conductor, Josephson current "wakes-up" "clock self" and initiates self-organization process. Therefore alarm clock is in question. Besides clocks and alarm clocks one can build pattern recognizers and novelty detectors and these circuits could serve as building bricks of complicated Boolean circuits controlling the functioning of living systems and also morphogenesis. Potential differences between gene space-time sheets and some larger space-time sheets, such as growing organs serving as the controllers of the gene expresion, would act as transcription factors in the sense that suitable input supra-current would "wake-up" gene self and activate self-organization process leading to gene expression.

Some examples are in order to show that this idea might have relevance.

1. The replication of the cell is an extremely complicated process but could be understood as quantum self-organization process leading to final state patter which only very mildly depends on the initial state. This process must be initiated by a "wake-up" self representing perhaps the cell. The alarm clocks must now be contained to the membrane surrounding the cell nucleus and probably also to the cell membrane since the cell membrane is known to be coupled to the division process of the cell nucleus, too [I10]. The reference currents are generated, when the new cell is born. The process leading to the replication of the cell involves a reduction in the density of super conducting charge carriers in the critical region and this could initiate the replication of the cell. This is achieved if Josephson currents run away from certain region of the membrane of the cell nucleus implying depletion of charge carriers.

2. The generation of completely new spatial structures during the morphogenesis is second extremely complicated process which should be understandable in terms of quantum self-organization. An example is afforded by the generation of somites [A1], which later give rise to brain and spinal cord. The homogenous longitudinal cell mass divides in a phase transition like manner into somites with clock wise regularity and the number of the somites is a constant characteristic for the species in question [A1]. The catastrophe theoretic models proposed in [A1] are based on the assumption that the pulse triggering the formation of somites is coupled to a biological clock, so that the motion of the boundary between differentiated and undifferentiated cell mass alternately slows down or fastens up and implies the generation of discrete regions, where the formation of the somites takes place.

A qualitative TGD based description is provided by the alarm clock model:

1. There is certain biorhythm realized using Josephson junctions (rhythms (minute scale) of this kind have indeed been identified [A1]) at cell level.

2. Josephson currents flow between the cells belonging to the longitudinal cell mass and neighboring cells in transversal direction. Due to the presence of the cell level reference currents,

Josephson currents interfere destructively and variations in density of charge carriers are small.

3. There is slow dependence of the phase of the order parameter ψ along the linear cell mass implying a phase lag between the clocks.

4. Reference current dissipates gradually through phase slippages and when the time is ripe the amplitude of the Josephson current becomes large and makes the density of charge carriers small inside the longitudinal region. The formation of the somites begins since the stability criterion implies that the stable size of topological field quantum decreases.

5. Time regulation is achieved through the presence of the biological clock: nothing happens unless the phase of the clock is correct since Josephson current runs to a "wrong" direction.

6. The process begins from the cells, which were born first since the clocks associated with them were created first and propagates in the order, in which the cells were born. In fact, the spatial dependence of the phase of the order parameter might code this order. The spatial dependence of the phase means that the rate for the propagation of the somite formation varies with position and guarantees in this manner the formation of spatially separated structures (compare with clock wave front model of [A1]). The number of the somites is just the multiple of 2π: s that the phase of the order parameter increases along the longitudinal cell mass.

7.2 Could Vacuum Quantum Numbers Control Gene Expression Via Josephson Currents

Controlled and synchronized gene expression is the most fundamental aspect of morphogenesis and implies surprising determinism of the development. When developing organism achieves certain level of development, certain gene activates. This requires feedback mechanism from long length scales of size of order organ to the gene level. In standard physics, the most plausible mechanisms are chemical. Whether this is the case is an unanswered question yet. In any case, it is very difficult to imagine how chemical concentrations which carry purely local information, could code information about the size of the organ and how the evolution could have led to a chemical kinetics initiating gene expression for critical values of the various chemical concentrations. The notions of many-sheeted space-time and general hypothesis about bio-control and coordination based on a hierarchy of weakly coupled super conductors provides a fresh and more promising approach to this process. This hypothesis is discussed in detail in [K12, K13].

Many-sheeted space-time concept suggests hierarchies of biological alarm clocks whose ringing induces ringing of some clocks at a lower level of hierarchy so that finally the alarm clock waking-up and activating definite gene, rings. One mechanism causing the ringing would be a situation in which the potential difference associated with the Josephson junction becomes equal to the energy difference for states associated with either super conductor: cyclotron resonance, which seems to be crucial for brain functioning and EEG, is basic example of this. This could at DNA level lead to the activation of gene and start up of a self-organization process. One could imagine complicated circuits in which ringing would occur only provided all the required conditions are achieved.

The correlation of gene expression with the size of the growing organ could be achieved as follows. Topological field quanta are characterized by a handful of vacuum quantum numbers associated with the dependence of the phases of the two CP_2 complex coordinates ξ^i on space-time coordinates (Appendix). In particular, two frequency type quantum numbers emerges. If the potential difference corresponds to the difference of the vacuum frequencies ω_1 associated with the coupled super conductors and if ω_1 correlates with the size of the corresponding structures, the ringing of the clock occurs when the size difference is critical. If the first super conductor corresponds to some structure with a fixed size (say gene) and second super conductor corresponds to the growing organ, this mechanism could indeed initiate new kind of gene expression when the growing organ reaches critical size.

7.3 Early Morphogenesis Of Drosophila

During the last years the understanding about the regulation of morphogenesis has grown dramatically [I11, I12]. For instance, in case of Drosophila (fruit fly) surprisingly detailed knowledge about the regulatory cascade occurring during embryogenesis exists. It is known that genetic program with 6 hierarchy levels is in action.

1. The so called maternal factors act before the onset of cellularization and lead to the segmentation of embryo. This cascade begins with the diffusion of transcription products of maternal-effect (coordinate-) genes from the anterior and posterior poles of the embryo during oögenesis. These genes define anterior-posterior polarization.

2. Maternal factors control the spatial pattern for the transcription of gap genes which are expressed in domains along the anterior to posterior axis of the embryo. Gap genes divide embryo to anterior, middle and posterior parts.

3. Gap genes regulate each other and the next set of genes in the hierarchy, pair-rule genes. These are expressed in 7 stripes of cells corresponding to every other segment.

4. At the next level of hierarchy are segment polarity genes, many of which are expressed in 14 segmentally repeated stripes. Segment polarity genes include also proteins other than transcription factors (i.e. secreted signalling molecules, receptors, kinases, etc.) and they mediate interactions between cells.

5. During cellular phase (blastula) Hox genes controlling the formation of various body parts are expressed. The lowest level of the hierarchy is represented by tissue specific genes.

The expression domains of these genes are indeed two-dimensional slices in accordance with the idea that genes can be regarded as obtained by compressing the expression domain of gene to DNA thread. This principle might be realized quite generally in the sense that the expression domains of all genes expressed inside particular body part are also slices formed as join along boundaries/flux tube condensates of fundamental expression domains whose size corresponds to some p-adic length scale.

7.4 Hox Genes

The discovery of the universality of so called hox (homeo box) genes has been one of the great discoveries in genetics during last few years years [I11, I12]. Surprisingly, animal species with widely different morphologies (hydra, flies, leeches, mouses and humans) seem to obey very similar body plan coded by Hox genes. Hox genes contain highly conserved nucleotide sequence called homeobox coding protein consisting of 61 amino-acids (which corresponds to DNA length of about 61 nanometers which is slightly below $L(157)$). Hox genes are also known as "selector" genes because their expression within a region of the embryo causes its cells to select a particular route of morphogenetic development determined by detailed nucleotide content of Hox gene. Hox genes function by coding what are called transcription factor proteins: these proteins bind to and activate all downstream genes necessary for the production of, say leg. Hence Hox gene for leg functions like a main program for the development of leg. Obviously, Hox genes provide a possibility to test and develop TGD based ideas about the coding and decoding of the morphology in terms of the many-sheeted DNA.

7.5 Evolution Of Hox Genes

The conserved nature of the homeobox sequence indicates that all Hox genes are homologous, having arisen by divergence from a common ancestral gene. Note however that the control parts of genetic program (promoter and operator regions at which various activating and repressing proteins bind) involved with Hox genes seem to vary widely. Second basic feature of Hox genes is their clustering. The development of this clustering has been studied and it has been found that all animals species must have inherited their Hox genes from a common ancestor.

1. Already plants possess single homeobox unit in their Hox gene. Doubling of the hox gene meant the emergence of a primitive head-body structure and thus of primitive animals (hydra, which are freshwater polypes, represent one studied example). Next step was doubling of this gene pair, which lead to the formation of nematode worms. The further doublings lead gradually to more complicated animals. For instance, beetle possesses 8 Hox genes, amphioxus, which is almost vertebrate, has 12 Hox genes.

2. The next step in the evolution was the doubling of the entire chromosome containing Hox cluster. The next doubling led to vertebrates with four Hox clusters located in four chromosomes. Each cluster contains a subset of 13 nonhomologous Hox genes. These clusters are labelled as Hox-A, Hox-B, Hox-C, Hox-D. n: th gene in, say cluster A, is denoted by Hox-A-n, $n = 1, .., 13$. The homologous Hox genes in various chromosomes, Hox-A-n, ..., Hox-D-n, whose number is never larger than four, form 13 groups called prologues. The maximum number of Hox genes is 52 but some of genes are missing. Each vertebrate has its own Hox-bode telling which Hox genes are absent and which are not. For instance, human 39 Hox genes.

Doubling of Hox clusters led to a more flexible gene expression since the conditions associated with gene statements involve protein outputs from all four Hox clusters. For instance, the condition in genetic program could have the form *IF [Hox-A-n or Hox-B-n] then...* This implies that the mutation of Hox gene in single cluster, which otherwise might be lethal, need not have any dramatic consequences. Doubling of the Hox cluster leads also to a multiplication of possible Hox patterns in breeding by a factor of 4 and hence to the large variation of the progeny, which is selective advantage. It is known that doublings of Hox clusters where followed by an emergence of large number of Hox-codes and only some of these codes have survived.

The presence of 4 Hox clusters in vertebrates could relate to four different tissue types corresponding to epithelial, connective, muscular and nervous tissues. Certainly this correspondence is not simply tissue-type $\leftrightarrow$ Hox cluster although there is some evidence that given Hox cluster dominates the expression of given tissue type under normal circumstances. For animals with 2 or 1 Hox clusters this kind of correspondence cannot hold true.

7.6 Characteristic Features Of Hox Genes

The characteristic features related to the expression of Hox genes give clues as how many-sheeted DNA is expressed.

7.6.1 Posterior Hox genes dominate over anterior Hox genes

Posterior Hox genes dominate over more anterior genes. For instance, if posterior Hox gene is removed to more anterior position the phenotype is more posterior. A simple illustrative example is based on a toy model of insect based on 3 Hox genes. The protein product of the Hox 1 gene instructs the formation of the head. The cells in the middle of body respond to both Hox1 and Hox2 genes. The protein product of Hox gene 2 is however believed to instruct the cells in this region not to not respond to Hox gene 1 so that Hox 2 genes determine the resulting structure. In the similar manner Hox 3 gene dominates over Hox 1 and Hox 2 genes in the tail part of the fictitious insect. Actually this picture is oversimplified but gives a good grasp of the idea.

This corresponds to a simple recursive control structure in genetic program.
P(n-1): IF protein coded by n: th Hox gene is the highest Hox protein present then DO B(n) ELSE DO P(n+1).
B(n): Form the n: th body part.

The problem is to understand how this control structure is realized. If Hox gene corresponds to program B(n) and does not serve as repressor of lower level Hox gene expression then it would seem that control structure must involve some "new genetics".

1. Hox genes select genetic programs. The mechanism of selection is based on attachment of the protein products to the control regions of corresponding genes could be such that gene program n is activated only if Hox proteins Hox-m $m \leq n$ are attached to the gene program.

2. One possibility involving introns in essential manner is that the activation of n: th Hox gene deactivates lower level Hox genes automatically. One possibility is that activation corresponds to the change of roles of introns and exons in Hox gene. Before activation Hox-n gene would be in "abnormal" state so that its intron parts would code protein. This protein could enhance the expression of the lower level Hox genes. After the activation this protein product would be absent and lower level Hox expression would not be enhanced anymore.

7.6.2 Hox expression domains are co-linear with the gene ordering inside Hox cluster

The development of embryo occurs in anterior to posterior (head to tail) direction. In all species hitherto examined, there is a strict correspondence between the ordering of the Hox genes inside clusters and the anterior boundaries of the expression domains along the head-tail axis of the developing embryo. Presence of the anterior boundary means that gene is not expressed above this boundary. The anterior boundaries of homologous genes in different clusters are same.

7.6.3 Establishment of Hox gene expression patterns in vertebrates

It is known that in vertebrates Hox gene expression patterns in developing embryo are established by waves propagating in posterior-anterior direction (tail to head). This means that there is also temporal co-linearity. Wave proceeds and ultimately stops at the anterior boundary of the expression domain. These anterior boundaries are characteristics of Hox genes and their ordering is the same as the ordering of Hox genes.

It is known that this pattern is not due to a forward spreading of cells. The presence of chemical signalling during the wave propagation is also excluded by the result of an experiment in which embryo was transversally sectioned [I11]. What happened that the wave propagated through the sectioning. Of course, it could be that chemical concentration gradient has been formed in earlier stage of development.

7.7 TGD Based Model For Hox Genes

The attempt to understand the known facts about Hox genes and their expression provides strong constraints on the general TGD based model of many-sheeted DNA and the results might be generalized to build more general models of gene expression during morphogenesis.

7.7.1 Hox cluster as a set of many-sheeted Hox genes?

Hox genes define division of developing embryo to slices of roughly same size. This suggests strongly that the largest space-time sheets associated with Hox genes correspond to the same p-adic prime $p \simeq 2^k$, k prime or power of prime. It could be that Hox genes are glued on same space-time sheet to form Hox cluster. Dominance of the posterior genes over anterior genes and spatial and temporal co-linearity reflect some kind of hierarchical ordering of Hox genes. This hierarchical ordering does not however seem to reflect the hierarchy of space-time sheets but gene program hierarchy.

7.7.2 Does activation of Hox gene involve a phase transition?

In vertebrates the activation of Hox genes occurs as a wave propagating from tail to head. A possible TGD based identification for the wave is as a phase transition leading to an expansion of a new space-time sheet associated with the many-sheeted Hox gene propagating in head to tail direction. This phase transition is determined solely by the internal state of the genome in given embryonic cell. In this phase transition the DNA space-time sheets would expand and join to form larger space-time sheet determining the size of the "imaginary disk" associated with the organ in question.

In Drosophila the length of DNA per single Hox gene plus its control region is about 10^{-5} meters. This suggests that the maximal space-time sheets of Hox genes correspond to $k = 173$, and that in activation this space-time sheet grows to its actual size. Of course, this growth could have occurred already in the segmentation stage. This would imply that the cells containing

Hox genes are glued together by flux tubes to form larger space-time sheets which later grow to space-time sheets corresponding to various organs.

7.7.3 How to understand basic facts about Hox gene expression?

There are several aspects involved with Hox gene expression which one should understand.

1. Co-linearity.
   Somehow the activation of n: th Hox gene leads to the activation of n+1: th Hox gene independently of what the n+1: th gene is. One possibility is that there is transcription factor gradient along the Hox gene which grows with time and gradually activates Hox genes in linear order. Perhaps enhancer is in question. Second possibility is that many-sheeted nature of Hox genes is crucially involved. Suppose that activation of Hox gene involves expansion of the largest space-time sheet associated with Hox gene. If the expansion of the space-time sheet of n: th Hox gene is necessary condition for the expansion of n+1: th space-time sheet, co-linearity follows automatically. This looks natural if the expansion of the Hox gene space-time sheet proceeds slowly along the Hox cluster so that expanding space-time sheet of n: th Hox gene is glued with that of n+1: th Hox gene

   This picture explains automatically also why the relocation of the Hox gene to a more anterior position makes Hox expression more posterior. The anterior boundaries in Hox clusters located in different chromosomes are the same. This suggest that the control for the beginning of the Hox expression program is associated with the space-time sheet representing entire embryo. Control could involve classical field affecting acting via enhancer or silencer type transcription factors whose effect is known to be not purely chemical.

2. Temporal co-linearity.
   This is present in vertebrates but not in Drosophila. Chemical signalling between cells cannot explain temporal co-linearity and it must be due to the independent development of the cell genomes. The simplest explanation is that there is lag in the activation time for the genetic program activating Hox genes increasing monotonically as a function of the distance along the anterior-posterior axis. This dependence could be generated by a gradient of corresponding transcription factor concentration, perhaps created already in maternal period. This gradient should be generated by a DNA sequence located at the $5'$ end of the Hox cluster so that transcription factor in question must be repressor: activation takes place when the concentration is subcritical.

3. The presence of the anterior boundaries.
   The presence of the anterior boundaries is most naturally due to determination of cells occurred before the activation of Hox genes. Segmentation genes indeed force various segments to different branches of genetic program and it is quite plausible that the activation of the Hox genes depends on segment. What is needed that the Hox genes above anterior boundary are too far from criticality for the activation to occur. It is also possible that cells in anterior end of the embryo generate repressor concentration which gradually grows and shifts anterior boundary in posterior direction during morphogenesis. If the expansion of space-time sheets associated with Hox genes is responsible for activation, this expansion should stop at n: th gene in n: th expression domain. It is not clear why this should occur without the proposed mechanism.

4. Posterior prevalence.
   An explanation in terms of the control mechanisms of gene programs activated by Hox genes has been already considered. Second explanation for posterior prevalence is that the expansion of the space-time sheet associated with Hox genes to a single space-time sheet common to first n Hox genes somehow represses the expression of all Hox genes expect n: th one. Perhaps only the gene on the boundary of this space-time can be active.

7.7.4 Quantum model for the expression of Hox genes

There are too much unknown factors to allow a construction of a detailed quantum model for the situation. What however seems clear that the differences of the vacuum frequencies representing

potential differences over Josephson junctions should effectively appear as transcription factors and control parameters. A quantum model possibly catching some aspects of Hox gene expression might involve at least the following assumptions.

1. Assume that genes along chromosome are characterized by vacuum frequencies ω_i. This frequency does not depend on gene but only on its position in the chromosome. This assumption guarantees spatial co-linearity. The dependence of the gene's vacuum frequency on the position of gene along the chromosome could be understood if chromosome forms linear join along boundaries/flux tube condensate with Josephson junctions connecting subsequent gene space-time sheets. This would mean that there is electric field along the chromosome. These Josephson junctions could be involved with the activation of the gene by the control DNA section preceding it. Also the effect of the enhancers and silencers might involve resonant Josephson current between control section of DNA and gene leading to the wake-up of the gene.

2. The vacuum frequency Ω characterizing the size of the growing organism increases with time and Ω changes in a phase transition like manner step by step. Temporal co-linearity can be understood if there is some (possibly phase) gradient along the growing organism implying that the phase transition leading to the increase of Ω proceeds from tail to head. The gradient could act like a concentration of a chemical suppressor decreasing in head-to tail direction and established already before the Hox gene expression started.

3. The space-time sheet of the organism and gene space-time sheets form weakly coupled super conductors and potential differences over the Josephson junctions serve as transcription factors. When the frequency $\Omega - \omega_i$, which corresponds to potential difference eV, equals to critical frequency, resonant currents are generated waking-up i: th gene and activating it. Given gene is active only during the time interval when $\Omega - \omega_i$ is critical. With the assumed dependence of ω_i on the position of gene, this implies spatial co-linearity and posterior prevalence.

REFERENCES

Mathematics

[A1] Zeeman EC. *Catastrophe Theory.* Addison-Wessley Publishing Company, 1977.

Condensed Matter Physics

[D1] Liquid crystals on line. Available at: `http://www.lcionline.net/`.

Biology

[I1] Harpalinae. Available at: `http://phylogeny.arizona.edu/tree/carabidae/harpalinae.html`.

[I2] Introns. Available at: `http://en.wikipedia.org/wiki/Introns`.

[I3] Kiger JAJr Ayuala FJ. *Modern Genetics.* Benjamin Cummings, 1984.

[I4] Benoit et al. Retrohoming of a Bacterial Group II Intron: Mobility via Complete Reverse Splicing, Independent of Homologous DNA Recombination. *Cell*, 94:456–462, 1998.

[I5] Tycowski T K et al. A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, 379:464–466, 1996.

[I6] Yang J et al. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature*, 1996.

[I7] Amato I. DNA Shows Unexplained Patterns. *Science*, page 747, 1992.

[I8] Ho M-W. *The Rainbow and the Worm*. World Scientific, Singapore, 1993.

[I9] Ho M-W. Coherent Energy, Liquid Crystallinity and Acupuncture. Available at: http://tinyurl.com/q3qztcn, 1994.

[I10] Volkenstein MV. *Biophysics*. Mir Publishers, Moscow, 1983.

[I11] Gaunt S. Hox Genes: Regulators of Animal Design. Available at: http://www.bi.bbsrc.ac.uk/WORLD/Sci4Alll/Gaunt/Gaunt2.html, 1999.

[I12] Brook W. Genetic control of segmentation in Drosophila: The maternal legacy, 1999.

Neuroscience and Consciousness

[J1] Yarrow D. Spin the tale of the dragon. Available at: http://www.ratical.org/reatvlle/RofD2.html, 1990.

[J2] Wu M Hu H. Action Potential Modulation of Neural Spin Networks Suggests Possible Role of Spin. *NeuroQuantology* . Available at: http://cogprints.org/3458/1/SpinRole.pdf, 4:309–317, 2004.

[J3] Hitt J. This is Your Brain on God. *Wired* . Available at: http://www.wired.com/wired/archive/7.11/persinger_pr.html, 1999.

Books related to TGD

[K1] Pitkänen M. Construction of Quantum Theory: M-matrix. In *Towards M-Matrix*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdquantum/tgdquantum.html#towards, 2006.

[K2] Pitkänen M. Construction of Quantum Theory: Symmetries. In *Towards M-Matrix*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdquantum/tgdquantum.html#quthe, 2006.

[K3] Pitkänen M. Dark Matter Hierarchy and Hierarchy of EEGs. In *TGD and EEG*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdeeg/tgdeeg.html#eegdark, 2006.

[K4] Pitkänen M. DNA as Topological Quantum Computer. In *Genes and Memes*. Onlinebook. Available at: http://tgdtheory.fi/public_html/genememe/genememe.html#dnatqc, 2006.

[K5] Pitkänen M. Does TGD Predict the Spectrum of Planck Constants? In *Hyper-finite Factors and Dark Matter Hierarchy*. Onlinebook. Available at: http://tgdtheory.fi/public_html/neuplanck/neuplanck.html#Planck, 2006.

[K6] Pitkänen M. General Theory of Qualia. In *Bio-Systems as Conscious Holograms*. Onlinebook. Available at: http://tgdtheory.fi/public_html/hologram/hologram.html#qualia, 2006.

[K7] Pitkänen M. Genes and Memes. In *Genes and Memes*. Onlinebook. Available at: http://tgdtheory.fi/public_html/genememe/genememe.html#genememec, 2006.

[K8] Pitkänen M. Homeopathy in Many-Sheeted Space-Time. In *Bio-Systems as Conscious Holograms*. Onlinebook. Available at: http://tgdtheory.fi/public_html/hologram/hologram.html#homeoc, 2006.

[K9] Pitkänen M. Many-Sheeted DNA. In *Genes and Memes*. Onlinebook. Available at: http://tgdtheory.fi/public_html/genememe/genememe.html#genecodec, 2006.

[K10] Pitkänen M. Negentropy Maximization Principle. In *TGD Inspired Theory of Consciousness*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdconsc/tgdconsc.html#nmpc, 2006.

[K11] Pitkänen M. Nuclear String Hypothesis. In *Hyper-finite Factors and Dark Matter Hierarchy*. Onlinebook. Available at: http://tgdtheory.fi/public_html/neuplanck/neuplanck.html#nuclstring, 2006.

[K12] Pitkänen M. Quantum Control and Coordination in Bio-systems: Part I. In *Bio-Systems as Self-Organizing Quantum Systems*. Onlinebook. Available at: http://tgdtheory.fi/public_html/bioselforg/bioselforg.html#qcococI, 2006.

[K13] Pitkänen M. Quantum Control and Coordination in Bio-Systems: Part II. In *Bio-Systems as Self-Organizing Quantum Systems*. Onlinebook. Available at: http://tgdtheory.fi/public_html/bioselforg/bioselforg.html#qcococII, 2006.

[K14] Pitkänen M. Topological Quantum Computation in TGD Universe. In *Genes and Memes*. Onlinebook. Available at: http://tgdtheory.fi/public_html/genememe/genememe.html#tqc, 2006.

[K15] Pitkänen M. Was von Neumann Right After All? In *Hyper-finite Factors and Dark Matter Hierarchy*. Onlinebook. Available at: http://tgdtheory.fi/public_html/neuplanck/neuplanck.html#vNeumann, 2006.

[K16] Pitkänen M. WCW Spinor Structure. In *Quantum Physics as Infinite-Dimensional Geometry*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdgeom/tgdgeom.html#cspin, 2006.

[K17] Pitkänen M. Are dark photons behind biophotons. In *TGD based view about living matter and remote mental interactions*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdlian/tgdlian.html#biophotonslian, 2013.

[K18] Pitkänen M. Comments on the recent experiments by the group of Michael Persinger. In *TGD based view about living matter and remote mental interactions*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdlian/tgdlian.html#persconsc, 2013.

[K19] Pitkänen M. Comparison of TGD inspired theory of consciousness with some other theories of consciousness. In *TGD based view about living matter and remote mental interactions*. Onlinebook. Available at: http://tgdtheory.fi/public_html/tgdlian/tgdlian.html#consccomparison, 2013.

Articles about TGD

[L1] Pitkänen M. Further Progress in Nuclear String Hypothesis. Available at: http://tgdtheory.fi/articles/nuclstring.pdf, 2007.

[L2] Pitkänen M. Quantum Model for the Direct Currents of Becker. Available at: http://tgdtheory.fi/public_html/articles/DCbio.pdf, 2012.

[L3] Pitkänen M. Basic Mechanisms associated with magnetic body.Available at: http://www.tgdtheory.fi/webCMAPs/BasicMechanismsassociatedwithmagneticbody.html. 2014.

[L4] Pitkänen M. CMAP representations about TGD. Available at: http://www.tgdtheory.fi/cmaphtml.html, 2014.

[L5] Pitkänen M. CMAP representations about TGD, and TGD inspired theory of consciousness and quantum biology. Available at: http://www.tgdtheory.fi/tgdglossary.pdf, 2014.

[L6] Pitkänen M. Dark proton strings and genetic code.Available at: http://www.tgdtheory.fi/webCMAPs/Darkprotonstringsandgeneticcode.html. 2014.

[L7] Pitkänen M. Genes and memes.Available at: http://www.tgdtheory.fi/webCMAPs/Genesandmemes.html. 2014.

[L8] Pitkänen M. Magnetic body.Available at: http://www.tgdtheory.fi/webCMAPs/Magneticbody.html. 2014.

[L9] Pitkänen M. Nuclear string model.Available at: http://www.tgdtheory.fi/webCMAPs/Nuclearstringmodel.html. 2014.

[L10] Pitkänen M. Origin of genetic code.Available at: http://www.tgdtheory.fi/webCMAPs/Originofgeneticcode.html. 2014.

[L11] Pitkänen M. Pollack's observations.Available at: http://www.tgdtheory.fi/webCMAPs/Pollack'sobservations.html. 2014.