Subcellular patterns of information processing

Nikita Tirjatkin

Structural and functional features of the cell are determined by information stored in DNA. This information is represented by a limited set of genes, a genome. Each gene can be expressed individually to be fully converted into corresponding element of the cell structure or function. During gene expression, the information processing typically involves DNA transcription, RNA translation, and catalysis. This sequence of chemical reactions can be called a gene expression network, abbreviated GEN. Within the cell, GEN is an universal pattern of information processing. It is essentially four-dimensional. From this perspective, the cell can be considered as a highly regular composition of interacting GENs, a GENome. The opportunity to recognize an universal pattern of information processing in the sequence of well-known reactions has been completely overlooked. Here, I draw attention to this pattern and show that its implication yields a powerful conceptual framework suited very well to strongly integrate known subcellular phenomena and reveal their novel emergent features.

From the information processing perspective, all reactions within the cell fall into three categories: DNA transcription, RNA translation, and catalysis.

During DNA transcription, information is converted from DNA into RNA form. In this process, the DNA serves as a template for the synthesis of a linear heteromers of ribonucleotides. This is possible because any ribonucleotides are allowed to be attached to the deoxyribonucleotides by means of base pair. In general, the ribonucleotides differ very slightly from the deoxyribonucleotides. Their sugar, the ribose, is quite equal to the deoxyribose. Although there is no ribonucleotide with the thymine as a base, there is one with a uracil that resembles the thymine not only structurally but also functionally in the forming a complementary pair with the adenine. Consequently, the adenine-, guanine-, cytosine- or uracil-carrying ribonucleotide can attach only to the thymine-, cytosine-, guanine- or adenine-carrying deoxyribonucleotide respectively. This specificity plays a key role in DNA transcription, where one strand separates from another and exposes a particular sequence of deoxyribonucleotides serving as a template by guiding the synthesis of a linear heteromer of ribonucleotides which sequence is complementary to the deoxyribonucleotide sequence exposed. Thereafter this heteromer of ribonucleotides separates from the DNA strand as a single-stranded polyribonucleotide, a ribonucleic acid, abbreviated RNA. DNA transcription may occur in different regions of the DNA molecule at the same time and may be repeated in the same region many times yielding lots and lots of different RNA molecules as end products. All RNA molecules produced from the same region of DNA molecule are always identical. The DNA molecule persists DNA transcription and remains unexhausted.

During RNA translation, information is converted from RNA into polypeptide form. In this process, the RNA molecule serves as a template for the synthesis of a linear heteromers of amino acids, or amino acid chains. The RNA molecule to be translated, referred to as a messenger RNA, abbreviated mRNA, can be considered not only as a

sequence of a single ribonucleotides but also as a sequence of a ribonucleotide triplets. The maximal number of possible combinations of four single ribonucleotides in one triplet is $4^3 = 64$. Thus 64 triplets can be distinguished. Two triplets can link together by means of base pairing to form a triplet pair. Because of the base complementarity, the triplet pairs can be considered also as complementary in the sense that if one member of the triplet pair is specified the other is also specified. This triplet pair complementarity makes RNA translation possible. The certain amino acid attaches an adaptor RNA molecule, referred to as a transfer RNA, abbreviated tRNA, that contains one specific triplet providing this tRNA the potential to specifically recognize and attach the complementary triplet on mRNA. By orderly forthcoming attachment of tRNAs to mRNA by triplets pairing, corresponding amino acids come enough close together to be interconnected by means of a peptide bonds to form an amino acid chain, referred therefore to as a polypeptide chain or, more familiar, a polypeptide. Thus, the mRNA sequence can guide the synthesis of the polypeptide that is considered to be complementary to this mRNA in the sense that if the sequence of triplets on the mRNA is specified the sequence of amino acids on the polypeptide is specified as well. So, because the sequence of a single ribonucleotides predetermines the sequence of triplets on the mRNA, it predetermines also the sequence of a single amino acids on the polypeptide and, therefore, all polypeptides produced on the same mRNA must be identical. The combination of a single ribonucleotides in the triplet is usually called a codon, because it specifies amino acid, i.e. specific triplet codes for specific amino acid, and the set of 64 possible codons is therefore referred to as a triplet code. The triplet code itself is however termed degenerate because it contains redundancies in the sense that most amino acids are encoded by more than one codon. Moreover, any codons do not specify amino acids but constitute stop signals that terminates RNA translation. Thus, the set of peptidogenic amino acids in cellular world is restricted to just 20 member. The triplet code has been highly conserved during evolution. The meaning of each codon is the same in virtually all present-day cells. There are only few minor exceptions in which some few codons have deviate meanings. The universality of the triplet code provides a strong suggestion that the cellular world on the Earth evolved only once and all present-day cells are descendants of a single primordial cell. As for the way how the triplet code might be arisen, it remains still a theme of speculations. The RNA molecule persists RNA translation without to be exhausted.

During catalysis, information is converted from catalyst into metabolite form. In this process, the catalyst serves as a template for the reaction that otherwise could occur too slowly for cell to live. The catalyst does its job of catalysis by physically grappling with one or more substrate molecules and interacting with them to make or break chemical bonds. The catalyst is usually very specific for the chemical reaction it catalyzes, and the specificity lies in a sophisticated configuration of atoms at one or more active sites of catalyst. Only restricted set of substrate molecules can recognize this configuration and bind it. In catalysts, this binding causes a conformational shift that promotes the reaction in any way. Thereafter, the catalyst releases reaction products, acquires its original conformation and is available for catalysis anew. Thus, the catalyst persists catalysis without to be exhausted.

Generally, within the cell, the DNA serves as a template in the synthesis of RNAs by DNA transcription. DNA transcription reactions produce large numbers of various

RNAs most of them are mRNAs that serve as a templates in synthesis of a polypeptides by RNA translation. RNA translation reactions yield especially large numbers of different polypeptides most of them are building blocks for enzymes that serve as a catalysts providing the cell with opportunity to tame a bewildering variety of different chemical reactions occurring in molecular world, consolidate and even incorporate most of them into its own network. In this way, the information stored in DNA becomes expressed.

Thus, from the information processing perspective, the cell can be considered as a DNA expression network.

It is important to note that the information is stored in the DNA as a limited set of genes, a genome. Each gene can be expressed individually to be fully converted into corresponding element of the cell structure or function. For each gene, its own sequence of DNA transcription, RNA translation, and catalysis can be determined. This sequence of chemical reactions can be called a gene expression network, abbreviated GEN.

In some GENs, however, this sequence can be restricted. So, in any GENs, the end products are polypeptides functioning always as substrate molecules and never as catalysts. There are also GENs which end products are RNAs that never become translated into polypeptides, but function always at the level of RNA as substrate molecules. On the other hand, the sequence of reactions in some GEN extends if the products of DNA transcription or RNA translation undergo the post-transcriptional or post-translational processing respectively. This additional processing involves both folding to acquire mature three-dimensional conformation and modification of selected ribonucleotides or amino acid to modulate structural and functional properties of mature RNA or polypeptide molecule. Often, some ribonucleotide or amino acid sequences are to be removed from the primary RNA or polypeptide. Particularly, single primary RNA or polypeptide is to be cut up into separate molecules or, on the contrary, some primary RNAs or polypeptides are to be joined together to form single molecule. Generally, the same gene can undergo expression many times thus yielding a large amount of identical end products.

Despite of differences in details, it is obvious that the GEN is an universal pattern of information processing in the cell. It is essentially four-dimensional. Moreover, the spatio-temporal organization of the GEN is characterized by the strong directionality of the so called flow of information.

All elements of the GEN are the best known subcellular phenomena. The directionality of information flow is just the best known component of the so called central dogma of biology. Nevertheless, the opportunity to recognize an universal pattern in the directed sequence of well-known reactions has been completely overlooked. Perhaps, the central dogma of biology was too dogmatic just in solving the central problem of biology.

In turn, the GENs are arranged into the whole DNA expression network where each GEN can be considered as an "adaptor" subnetwork which recognizes a particular piece of information on the DNA and transfers it into a corresponding fragment of the whole cell network. GENs do not work in isolation but receive inputs from each other and

from environment. In each GEN, there are many control buttons that determine when, where, and how much of particular gene end product is synthesized. Just the astonishing harmony of the whole DNA expression network reflects the strong integrity of information stored in the cell genome. From this perspective, the cell can be considered as a highly regular composition of interacting GENs, a GENome.

In GENome, GENs interact to maintain each other. During information processing in particular GEN, it is just the job of other GENs to provide necessary elements for gene expression machinery. It is reasonable to distinguish three main parts of this machinery according to the sequence of information processing reactions in GEN.

The DNA transcription machinery is necessary to convert information from DNA into RNA form. The DNA serves as a template in this process. In the DNA molecule overall, both strands can be used as templates, but in any one gene only one strand is used and in the same gene it is always the same strand. To cite the gene, however, just the deoxyribonucleotides sequence on its non-template strand is conventionally used. In particular GEN, the DNA is represented by a corresponding gene. First of all, the DNA transcription machinery involves a pool of four different ribonucleotides with the triphosphate as a phosphate group. Further constituents of the DNA transcription machinery are numerous complex molecules of various nature that work to polymerize ribonucleotides on the template. A RNA polymerase serves to find an appropriate site on DNA, to bind the DNA at this site, to temporally separate the two strands in the adjacent region, and to begin generating of an RNA molecule on one of the separated strands. Then, the RNA polymerase moves along the DNA, maintaining growing fork to expose the template strand, and catalyzes the elongation of RNA molecule by addition of the incoming free ribonucleotide to the 3' growing point. When the RNA polymerase recognizes specific region on DNA which signals for termination of DNA transcription, both RNA polymerase and primary RNA molecule are released from DNA. A multitude of molecules assist the RNA polymerase to perform these tasks. In some GENs, the DNA transcription machinery additionally involves constituents for post-transcriptional processing of primary RNA. Spliceosome containing several kinds of polypeptides and RNAs cuts primary RNA to remove some ribonucleotide sequences, introns, and then rejoin adjacent regions, exons. Ribozymes are RNAs with an enzymatic activity restricted to cleaving primary RNAs at specific locations. Some constituents of the DNA transcription machinery are universal, others are gene-specific.

The RNA translation machinery is necessary to convert information from RNA into polypeptide form. The mRNA serves as template in this process. First of all, the RNA translation machinery involves a pool of 20 different amino acids. Further, numerous complex molecules function to enable the polymerization of amino acids on the template. One prominent part of the RNA translation machinery is a pool of tRNAs. In this pool, 30 to 40 different tRNAs can be identified in some cells and 50 to 100 - in other cells. Different tRNAs show very similar three-dimensional conformation. This reflects their functional similarity: all have to be chemically linked to a particular amino acid and to attach a corresponding codon on mRNA so that the amino acid can be added to a growing polypeptide chain. Mature tRNA molecule resembles cloverleaf folded into L-like shape with two arms. The end of one arm has a specific triplet of ribonucleotides, termed the anticodon, which enables the tRNA to be specifically

attached to complementary codon on the mRNA. The end of other arm is a free 3' terminus of the tRNA which can be linked to amino acid. Another prominent part of RNA translation machinery is a pool of 20 aminoacyl-tRNA synthetases. Each tRNA molecule can be recognized by one and only one of the 20 aminoacyl-tRNA synthetases. Likewise, each of these enzymes links one and only one of the 20 amino acids to a particular tRNA, forming an aminoacyl-tRNA. The most prominent part of RNA translation machinery is a ribosome, an association of any few specific ribosomal RNAs, abbreviated rRNAs, with more than 100 different polypeptides. In ribosome, a large and a small subunit are usually distinguished. Physically moving along an mRNA molecule, the ribosome binds and brings together various accessory molecules necessary for polypeptide synthesis and catalyzes the assembly of amino acids into polypeptide chain. In some GENs, the RNA translation machinery additionally involves constituents for post-translational processing of primary polypeptide. Some constituents of the RNA translation machinery are universal, others are mRNA-specific. Although the mRNA persists information processing by RNA translation, it degrades soon or later.

The catalysis machinery is necessary to convert information from enzyme into metabolite form. The catalyst serves as a template in this process. First of all, the catalysis machinery involves a particular pool of substrate molecules. This pool is usually catalyst-specific. Further constituents of the catalysis machinery are numerous complex molecules of various nature that assist the catalysis. Although the catalyst persists information processing by catalysis, it degrades soon or later.

Collectively, GENs in GENome work to replicate the complete DNA.

During DNA replication, information is reproduced. The DNA serves as a template in this process. The structure of DNA molecule immediately suggests how it may be replicated: its strands must be first separated and then each strand can serve as a template guiding the synthesis of complementary strand according to the strict rules of base pairing. The separation of strands exposes their deoxyribonucleotides and thus makes each of them available to be recognized and attached by free complementary deoxyribonucleotide in the cell matrix. By orderly forthcoming attachment, newly added deoxyribonucleotides transform each single strand into duplex identical with original. So, qualitatively, the DNA replication is relatively simple chemical reaction. But, quantitatively, it is very complex because of a large amount of DNA in the cell. Therefore, the DNA is usually replicated not as a whole but rather in fragments of various lengths which then are to be joined together.

In the cell, the DNA can contain more than one DNA molecule. Each DNA molecule can contain more than one region where the DNA replication can begin. Such regions are called replication origins. At each replication origin, the separation of strands resembles unzipping of a zipper with forming of two growing forks that move outward in opposite directions. Each growing fork is a zone where two strands expose their deoxyribonucleotides. While the new deoxyribonucleotides can be added exclusively at the 3' growing point, the restoring of the duplex can proceeds continuously only on the strand which is exposed in the 5'-to-3' direction. This part of the growing fork is called a leading strand. In contrast, the restoring of the duplex on the other strand for which the

5'-to-3' direction is away from the growing fork can occur only in short stretches, called Okazaki fragments that are later to be joined together. This part of the growing fork is called a lagging strand. While at each replication origin two growing forks are formed, the new synthesis proceeds bidirectional. The duplexes that are being restored by each replication origin elongate and eventually join each other. When the DNA replication is complete, each original DNA molecule is replaced by its two identical copies called daughter DNA molecules. While each of them is half old and half new, this type of replication is termed semiconservative. After replication, each couple of daughter DNA molecules usually undergoes processing that involves their packing and separation from one another.

In the cell, the DNA is never free but rather in association with a large number of diverse polypeptides. Such composition is called a chromatin. In the chromatin, most of the polypeptides serve to compact DNA sufficiently to fit it inside the cell. In distinct areas, the chromatin shows different degree of condensation. In areas with lowest degree of chromatin condensation, DNA fragments are present usually in so called B form. In this form the stacked base pairs are regularly spaced 0.34 nm apart the helix axis and the helix makes a complete turn every 3.4 nm with about 10 pairs per turn. On the outside of DNA molecule in B-form, the spaces between the intertwined strands form two helical grooves of different widths referred to as the major groove and the minor groove. Consequently, part of each base is accessible from outside the helix to both small and large molecules that bind to the DNA by contacting chemical groups within the grooves. These two binding surfaces of the DNA molecule are used by different DNA-binding molecules. In addition to the major B form of DNA, A form and Z form have been described. First of all, the DNA replication machinery involves a pool of four different deoxyribonucleotides. Also a pool of four different ribonucleotides is needed. In both pools, nucleotides contain always the triphosphate as a phosphate group. Further, numerous complex molecules function to enable the polymerization of deoxyribonucleotides on the template. A helicase unwinds and unzips the duplex to form growing fork. The local unwinding and unzipping produce the torsional stress leading to supercoiling of region adjacent to the fork. A topoisomerase removes this supercoil. On the exposed DNA strand, a specialized RNA polymerase, also called a DNA primase, forms a short RNA strand which serve then as a primer that can be elongated by a DNA polymerase, thereby restoring the duplex. The primer is needed while the DNA polymerase is unable to initiate replication but rather to carry out a deoxyribonucleotide addition only. At each DNA fragment, the primer is removed and replaced by growth of the deoxyribonucleotide strand from the neighboring DNA fragment. A DNA ligase joins the adjacent fragments. Combined with the ability to discriminate against incorporation of a mismatched deoxyribonucleotide, the proofreading activity of DNA polymerase is sufficient to reduce the error frequency of DNA replication significantly. Additionally, numerous molecules act to remove mismatched or damaged deoxyribonucleotides, further ensuring correct DNA replication and maintaining integrity of genome.

During DNA replication, each original DNA molecule will be ultimately replaced by a couple of daughter DNA molecules. Consequently, when the DNA replication is complete, the pool of free deoxyribonucleotides is exhausted, DNA replication machinery disappears, and chromatin polypeptides enrich the matrix. The post-

replicative packing of each DNA molecule by chromatin polypeptides results in structure referred to as a chromatid. Respectively, the packing of each couple of daughter DNA molecules produces a couple of two chromatids held together. This structure as a whole is referred to as a chromosome, its both constituents are called sister chromatids. Finally, each chromosome undergoes separation and sister chromatids segregate.

Whereas gene and genome are notions that refer to how information is stored in the cell, GEN and GENome refer to how the gene and genome function. The cells differ greatly in number and assortment of genes in genome and, subsequently, in number and assortment of GENs in GENome.

In some genomes, particular gene can be present in multiple copies which are either spread throughout the genome or arranged in tandem arrays. Often, some genes in the genome can be considered as a family of related genes. Such families can comprise from any few to 100 or even 1000 members which also can be either spread throughout the genome or arranged in tandem arrays. In the cell, the DNA can be not only monogenomic but also digenomic. Digenomic DNA contains two non-identical genomes. Both monogenomic and digenomic cells can exhibit high level of ploidy which refer to the number of genomes without distinguishing between identical and non-identical.

Practically, the cell suits the GENome to specific subset of sources of mass, impulse, and energy to produce their usable forms essential for the cell life. These sources label the start points of different metabolic pathways which refer to the routes of mass, impulse, and energy processing. Initially separated, these pathways become more and more intricately interwoven, but then divide anew into many branches which end points are marked by waste products. In contrast to sources of mass, impulse, and energy, the information source is appointed not in environment but rather in the cell itself and, more concretely, just in the DNA. In the cell, the same DNA, i.e. the same DNA molecule or the same set of DNA molecules, undergoes not only DNA expression but also DNA replication and provides in this way the cell with the key mechanism to couple these two phenomena into single one. If information encoded in the sequence of deoxyribonucleotides of DNA will be executed by DNA expression, the DNA will be reproduced by DNA replication. In other words, the DNA promotes its own reproduction by DNA replication through execution of its own information by DNA expression. As a result, the DNA not only persists information processing without to be exhausted but also becomes transformed into two identical copies of itself during the cell life history.

From the perspective of mass, impulse, and energy processing, bewildering large number of reaction types is recognized in cells. There is a great deal of convenience by their integration into subcellular patterns. In contrast, GEN and GENome seem to be a natural basis for this integration. As for the patterns, such as metabolic pathways, signaling cascades, regulatory motifs, functional modules, etc., they rather disclose some reciprocal relations between GENs. Taking into account the arrangement of chemical reactions into GENs provides the reverse engineering of these patterns with common platform.

Although differing greatly in the number and assortment of sources of mass, impulse, and energy, the cells are very similar in preparation carefully of the appropriate pools of such molecules as 4 ribonucleotides and 20 amino acids. Virtually each cell contains GENs that need these precursors to execute information stored in corresponding gene. Different cells are also similar in preparation carefully of the appropriate pool of 4 deoxyribonucleotides. Virtually in each cell, GENome needs these precursors to reproduce information stored in corresponding genome.

Spatio-temporally, the cell appears as a more or less quickly changing matrix composed of a bewildering number of chemicals participating on reactions involved in GENome. This matrix is mostly a heterogeneous water solution of a large array of mono- and polymolecules, but one prominent constituent is a tiny membrane composed of amphipathic molecules. Whereas one part of an amphipathic molecule is hydrophilic, i.e. water soluble, another part is hydrophobic, i.e. water insoluble. Placed in water, such molecules aggregate, arranging their hydrophobic portions as much in contact with one another as possible and their hydrophilic portions in contact with water. The most appropriate three-dimensional configuration of such aggregate is a bilayered membrane closed to form a vessel-like shape. In the cell, most of amphipathic molecules are in this configuration separating the matrix into an interior and an exterior part respectively. In the literature, only the interior part with the membrane is mentioned to be a cell, and the exterior part is referred to as an extracellular matrix. Here is preferred, though, to mention under the cell a whole matrix with the DNA as a key constituent. Other chemicals become constituents of the cell matrix not occasionally depending solely on the conditions in environment but rather regularly depending entirely on the information stored in the DNA. In this respect, the whole cell matrix can be considered as ultimate derivative of DNA.

Although sharing common structural and functional properties that have been conserved throughout evolution of the cellular world, the cells have evolved a variety of differences. In this respect, two large cell groups are usually distinguished: prokaryotic cells and eukaryotic cells.

The prokaryotic cells are commonly small entities with relatively simple spatio-temporal organization.

They have mostly only one DNA molecule, but any few cell types do have 2 or more. In the DNA, the number of deoxyribonucleotide pairs ranges from about 0.6 million to 5 millions, an amount sufficient to contain 1000 to 4000 genes. The genes are arranged close together with little intergenic space. Introns are extremely rare. The DNA is usually a ring-formed molecule. Although it is attached to the cell membrane, its largest part lies in the central region of the interior matrix and is arranged together with numerous polypeptides in a dense clump called nucleoid. The circular DNA molecule has usually only one replication origin. In addition, numerous accessory DNA molecules, called plasmids, are distributed in the interior matrix. They depend on cellular machinery to be replicated or expressed and can not survive at all outside of the cell. Plasmids carry some genes that, although not essential for cell life, are extremely

useful to the host in any situation. In prokaryotic cell, there is only one type of RNA polymerase which does the job of synthesis of all types of RNA molecule.

The cell membrane is commonly a single simply-shaped vessel that undergoes very slight changes in cell life history. Some cell types have however a second cell membrane which lies concentric to the first and subdivides the exterior matrix additionally. A sphere, a rod, or a spiral are the most abundant shapes of cell membrane-formed vessel which linear dimensions range from 1 to 10 μm. But, the diversity of shapes extends well beyond these basic types. In any few cell types, the cell membrane region associated with the DNA molecule can invaginate. In some other cell types, the cell membrane regions containing specialized sets of polypeptides are able not only to invaginate but also to pinch off completely, forming sealed vesicles which become suspended in the interior matrix. In cell types with two concentric cell membranes, sealed vesicles are constantly being discharged from outer cell membrane into environment and attack neighboring cells. In interior matrix, there are also numerous granular inclusions such as ribosomes and thylakoids. Thylakoids are flattened discs with light-sensitive pigment molecules.

In the exterior matrix, a rigid cell wall is commonly formed that is composed of complex assemblies of polypeptides and polysaccharides and is porous. In cell types with two concentric cell membranes, this wall is usually placed between them. The cell wall supports the cell to maintain its shape and provides it with additional mechanisms to protect key chemical reactions that occur preferentially in the interior matrix and at the cell membrane. Other prominent part of exterior matrix is a glycocalyx (capsule, gelatinous sheath or slime layer). Some prokaryotic cell types have flagella, which rotate like propellers to move cell through fluid medium. Fimbriae are short appendages that help cell attach to an appropriate surface.

The composition of metabolic pathways is relatively simple in prokaryotic cells, but, they show a broad variability in this respect. Some prokaryotic cells require only carbon dioxide as a carbon source. Certain prokaryotic cells use light as energy source. The others can oxidize various chemicals to obtain energy. Some prokaryotic cells can live only in the presence of oxygen, others only in the absence. Certain prokaryotic cells can live in the presence or absence of oxygen. Different metabolic pathways are often combined in the same cell and it can switch between them repeatedly. Some prokaryotic cells can also survive in environments that are considered as extremely hostile. They can inhabit extremely hot or cold habitats. They also can reside in extremely salt, acid, or alkaline surroundings.

The eukaryotic cells are generally much larger than prokaryotic cells and show relatively complex spatio-temporal organization.

They contain 1 to more than 50 long DNA molecules which are usually linear. On average, the DNA in a typical eukaryotic cell is about 1000 times greater in number of deoxyribonucleotide pairs than a DNA in a typical prokaryotic cell. Although a large amount of deoxyribonucleotide sequences in the DNA is non-coding, the average number of genes in a representative eukaryotic genome is still tenfold greater than in a

representative prokaryotic genome. The genes are distributed between different DNA molecules. Thus, each DNA molecule contains only the part of the genome.

The cell membrane is quite an elaborated system that changes significantly in cell life history.

In addition to intricately-shaped main cell vessel, which is 10 to 30 times larger in linear dimension and 1000 to 10,000 times greater in volume than a related vessel of typical prokaryotic cell, the cell membrane forms also an expanded set of various vesicles suspended in the interior matrix. Each vesicle has a lumen which content is considered to be topologically equivalent to the exterior matrix. The vesicles vary in dimensions and shape. Intensively communicating with one another, they compose an integral vesicular system.

The most prominent vesicle is a nuclear envelope. Its membrane makes up two concentric surfaces which lay close together with little space between them. Therefore, the nuclear envelope-forming vesicle looks like a double-membraned barrier subdividing the interior matrix into two functionally distinct compartments: a nucleus and a cytosol. The nucleus contains the DNA and is a principal site of DNA replication and DNA transcription. The eukaryotic cell has three different RNA polymerases, which specialize for synthesis of tRNAs, rRNAs, and mRNAs respectively. Sites where rRNAs join polypeptides to form ribosomes are called nucleoli. The cytosol is the site of RNA translation and of most of the cellular metabolism. The nuclear envelope is punctured at intervals by nuclear pores where membranes of two concentric surfaces fuse. The nuclear pores serve to actively transport molecules between nucleus and cytosol. The lumen of the nuclear envelope, referred to as a perinuclear space, continues into the labyrinthine lumen of the endoplasmic reticulum, a system of large flattened vesicles. The nuclear envelope and endoplasmic reticulum compose a part of vesicular system where molecules destined for secretion undergo processing.

Golgi apparatus, organized stacks of disc-like Golgi cisternae, is another part of vesicular system where molecules destined for secretion undergo sorting. Numerous small vesicles bud off from one area of vesicular system and fuse with another. In similar ways, the vesicular system interacts with the main cell vessel. Lysosomes are vesicles containing digestive enzymes. They fuse usually with vesicles containing food molecules. Vacuoles are vesicles to store water, sugars, salts, pigments, or toxins.

Most eukaryotic cell types also contain vesicles which are occupied by a bodies so resembling a prokaryotic cells that they are thought to originate from them. Indeed, there are numerous arguments to assume that the primordial eukaryotic host cell, which formed symbiotic association with any prokaryotic cells, assimilated them as so called mitochondria and plastids respectively. Although both mitochondria and plastids lost through time most of the genes originally presented in the genome of their precursors, they preserved however any part of their DNA and also the ability to self-double and self-divide. DNA of mitochondria and plastids is circular loop similar to prokaryotic DNA. As for the ways how the symbiotic form might be arisen, it is a matter of dispute. Whereas mitochondria can be found in most of eukaryotic cells, plastids are restricted to

any few cell types. Individual mitochondrion or plastid can contain many copies of its own DNA.

Ribosomes may be attached to endoplasmic reticulum on cytosol site or may lie free in cytosol. Sometimes, ribosomes combine into polyribosomes. Sites of endoplasmic reticulum with attached ribosomes are called rough endoplasmic reticulum. Primary polypeptides synthesized at ribosomes enter lumen of endoplasmic reticulum for post-translational processing. Mitochondria and plastids have also their own ribosomes to produce own polypeptides. These ribosomes resemble prokaryotic ribosomes.

Numerous polypeptides in the cell matrix are arranged into so called cytoskeleton giving the cell strength and rigidity, thereby helping to maintain cell shape. Other polypeptides are arranged into structures enabling both the movement of the cell as a whole and the movement of various structures within the cell matrix. The main cell vessel may have numerous short hair-like projections, cilia, that can move in an undulating fashion or few longer whip-like projections, flagella that move in whip-like fashion. Both cilia and flagella have similar construction, but differ from prokaryotic flagella.

In the exterior matrix of diverse eukaryotic cell types, more or less rigid cell wall is built. Also the glycocalyx is often present in exterior matrix. It may contain scales, spicules, spines, shells, sheaths, tests, thecae, or loricae which are often very complex structurally.

Eukaryotic cells show only a few variability according to the composition of metabolic pathways but it is usually very complex.

To describe a spatio-temporal organization of a single cell adequately, it is enough to outline how its spatial organization changes temporally.

While in the cell the DNA is produced only by DNA replication, it is reasonable to assume that the cell life history begins at the point where two newly produced sister cells halve the matrix inherited from the mother cell and each starts a self-dependent life. What the newborn cell has to do is just what its mother done: self-maintain as long as possible to self-double and then to self-divide. Because its DNA carries the same genome as the mother DNA, all abilities to fulfill this plan are inherited as well. To ensure that this plan will be fulfilled, all chemical reactions of the cell network must be well organized with respect to each other in space and time. This comprises also the precise logistics of a bewildering number of chemicals participating on these reactions and constituting the cell matrix.

In different cells, the spatio-temporal organization is differently complex.

In prokaryotic cells, the cell life history is short and simple.

In ideal environmental conditions, the prokaryotic cell needs only 30 minutes to self-double and to self-divide. Temporally, DNA expression and DNA replication seem to occur at the same time and proceed concurrently. Replication of single DNA molecule

begins at the single replication origin which is anchored to the cell membrane. Segregation of new DNA molecules appears to begin soon after starting of duplication of the replication origin, whereas the remainder of the DNA molecule awaits replication. Once DNA replication is complete, an assembly of new membrane and cell wall forms a septum, which eventually divides the cell in two. Because the origins of the two newly formed DNA molecules are anchored to different membrane sites, each daughter cell receives one DNA molecule. Some prokaryotic cells are capable to initiate the next round of DNA replication before the previous round is complete. This results in a cell with multiple bidirectional replication forks but only a single unduplicated terminus region. During this multifork replication, new replication origins soon after duplication undergo segregation in opposite directions.

In eukaryotic cells, the cell life history is much longer and more complex than in prokaryotic cells.

The duration of the cell life varies greatly from one cell type to another. Any few eukaryotic cells can grow and divide as quickly as prokaryotic cells, while the cell life of some other can last longer than a year. However, a typical eukaryotic cell needs usually 10 to 20 hours to self-double and to self-divide.

Temporally, the cell life history in an eukaryotic cell is traditionally divided into four major periods.

During the first period, G_1 phase, the newly born cell begins with DNA expression, cell matrix continuously grows in mass and volume, and organelles increase in number. The cell monitors conditions in its own matrix and in environment and, when the time is ripe, takes a decisive step that commits it to DNA replication. This period is considered to be a safety gap.

The second period, S phase, represents just the time of DNA replication. Polypeptides associated with DNA in chromatin are also produced. Those regions of chromatin that are least condensed and therefore most accessible to the replication machinery are replicated first, whereas the condensed chromatin tends to be replicated very late in S phase. Since eukaryotic DNA molecules are lineal, there is no place to produce the RNA primers needed to start the last Okazaki fragments at the very tips so that the strands become shorter after replication. To solve this end-replication problem, eukaryotic cells produce special enzyme, telomerase, which is able to restore the length of the strands. In contrast to prokaryotic cells, DNA expression and DNA replication in eukaryotic cells proceed concurrently only during a restricted period of cell life history.

The third period, G_2 phase, provides the next safety gap, allowing the cell to ensure that DNA replication is complete before it plunges into cell division.

During the fourth period, M phase, the cell division occurs. In a comparatively short M phase, the contents of cell matrix, which were doubled by activities of the preceding three periods, collectively called interphase, are to be exactly segregated into two daughter cells.

In process of cell division, two subprocesses are usually distinguished: mitosis and a cytokinesis.

The mitosis, also called karyokinesis, refers to the segregation of the reduplicated DNA, the cytokinesis is a division of the cell matrix as a whole. The mitosis is often considered as a culmination of the cell life history. With minor variations, the events in mitosis follow the same sequence in all eukaryotic cell types. Although they unfold continuously, the mitosis is conventionally divided into five subperiods.

During the first subperiod, prophase, the cell prepares itself for mitosis. In the nucleus, the chromatin rearranges to condense into chromosomes, each comprising two identical chromatids linked together at a multiple points along their length. In the cytosol, the cytoskeleton rearranges to build a specialized machinery, a spindle apparatus, which is able to capture chromosomes and segregate sister chromatids from one another.

During the second subperiod, prometaphase, the nuclear envelope breaks up into multiple small vesicles and chromosomes are released in the cytosol and captured by the spindle apparatus.

During the third subperiod, metaphase, the cell seems to pause until all chromosomes are aligned appropriately at the middle plane between the two poles of the spindle, poised for segregation.

During the forth subperiod, anaphase, the sister chromatids on each chromosome abruptly split apart and are pulled by the spindle apparatus to its opposite poles. Meanwhile, the cell itself stretches out and the spindle apparatus pushes its poles apart and elongates. Whereas the pulling moves the chromatids to segregate them into two equal subsets, one at each pole of the spindle, the pushing increases the distance between these subsets.

The fifth subperiod, telophase, is marked by a disappearance of the spindle apparatus and by a coalescence of small vesicles to form a nuclear envelope around each subset of chromatids establishing thus the two daughter nuclei. In each newly formed nucleus, the chromatin begins to decondense.

Thus, in contrast to prokaryotic cells, segregation of new DNA molecules in eukaryotic cells is temporally separated from DNA duplication.

The cytokinesis accompanies the mitosis, beginning usually in anaphase, continuing in telophase, but finishing later as the mitosis is complete. Just the cytokinesis, by which the cell is divided into two daughter cells, is traditionally viewed as the end of the M phase and the cell life history.

In each period or subperiod, the cell is in a specific stage. So, the cell life history can be considered as a sequence of these stages. The transition of the cell into each sequential stage is a precisely regulated process, overseen by numerous genes whose job is to ensure that this sequence is carried out correctly. There are numerous critical checkpoints, at which the transition into a particular stage of the cell life history can be

arrested and delayed, if the cell is hindered by unfavorable conditions to complete previous events without errors or even damaged by any agent. Arrests and delays at critical checkpoints provide time for cell to correct errors and repair damages. The critical checkpoints are points at which the cell life history may be regulated by extracellular signals.

Duration of each period or subperiod respectively varies considerably in different cells.

In some cells, there is no G_1 and G_2 phase and DNA replication and cell division occur very rapidly so that both S and M phase are very short. In contrast, other cells exit G_1 phase to enter a phase of replicative quiescence, a G_0 phase, where they become arrested unless called on to leave replicative quiescence by an appropriate extracellular signal. If they acquire no signal to do so, they may even cease S, G_2, and M phases altogether and proceed into the replicative senescence. During G_0 phase, metabolic activity of the cell may vary significantly.

During its life history, the cell often changes drastically structural and functional characteristics.

For example, the cell can alternate between the motile and sessile life styles or change repeatedly between the locomotion by means of pseudopodia and swimming by means of flagella. It can also alternate between the active and passive life style or even become a dormant spore or cyst for a long period of time.

The cell may divide either symmetrically producing two identical daughter cells or asymmetrically producing two non-identical daughter cells.

The asymmetric cell division involves unequal segregation of the cell fate determinants between the two daughter cells during cytokinesis. These determinants may specify cell fate either directly through intrinsic metabolic pathways or indirectly through interaction with extrinsic determinants but, in each case, the cell fate specification is always achieved by differential DNA expression. The DNA expression may be made differential at each reaction of the DNA expression network in many ways.

Generally, the life history of the single cell begins with one cell but ends with two.

The ability of the DNA molecule to be produced not only casually with a sequence depended mainly on chance but also causally with a sequence predetermined by a prototype make this molecule unique and suit it well for its specific role in origin and evolution of the cellular world.

Whereas in the molecular world the DNA molecule was merely a side-product of evolution and the DNA replication occurred to some extent occasionally and seldom, exactly the DNA molecule became the key substance and the DNA replication - the key reaction that both made the origin of the cellular world just possible. In molecular world on the early Earth, the same substance might be occasionally involved in more than one chemical reaction. As a result, a special network could evolve in which so coupled reactions promote each other. Likewise, the DNA was one time involved in a

sophisticated network of intimately interconnected chemical reactions and so promoted to successfully transform the DNA replication from an occasional event into a highly persistent one. As for the way by which this network might originate, it is still a matter of dispute. Most important, while involving the DNA replication, this network, once originating billions years ago, persists until now by progressive self-propagation and self-diversification.

Moreover, while progressively self-propagating and self-diversifying since its origin, this network tirelessly develops itself to only one super-network that is so huge that it is impossible to imagine and describe its spatio-temporal organization realistically. Temporally, this super-network stretches itself out through approximately 3 or 4 billions years. Spatially, first as small as a single mycoplasma, it is since billions of years as large as the whole biosphere of the Earth. The present-day biosphere is merely a visible top of iceberg in ocean of time. The ancient part of this gigantic life pattern leaves very scarce traces. In the known universe, it is the most complex structure that just represents a whole cellular world, and one part of the super-network, the cell, is often considered as an atom of the cellular word in the sense to be their smallest entity. As for the other respects, this "a-tom" is in reality a true auto-tom: the self-division is its essential feature. A self-dividing cell creates a bewildering diversity of life patterns in the cellular world.

In conclusion, the proposed conceptual framework based on notions of GEN and GENome seems to be suited very well to strongly integrate known subcellular phenomena and reveal their novel emergent features. Whereas gene and genome are notions that refer to how information is stored in the cell, GEN and GENome refer to how the gene and genome work. During information processing in particular GEN, it is just the job of other GENs to provide necessary elements for gene expression machinery. Collectively, GENs in GENome work to replicate the complete DNA so that the life history of the single cell begins with one cell but ends with two. Practically, the cell suits the GENome to specific subset of sources of mass, impulse, and energy to produce their usable forms essential for the cell life. Spatio-temporally, the cell appears as a more or less quickly changing matrix composed of a bewildering number of chemicals participating on reactions involved in GENome.

Information to article

Written **14 July 2005**.
Published online at *www.nikita-tirjatkin.de* **14 July 2005**.

© 2005 Nikita Tirjatkin
This is an Open Access Article. Unrestricted use, reproduction, and distribution are permitted in all media for any purpose, provided the original work is properly cited.