Creation Research Society Quarterly 2016. 53:19–41.

Cells as Information Processors

Part 2: Hardware Implementation

Dr. Royal Truman*

Abstract

Cells use many codes, which operate through abstract symbolism and formal rules. To instantiate the logic, dedicated decoding processors—hardware molecular machines—must interpret each kind of variable and the associated values. Decoders include ribosomes, DNA polymerases, RNA polymerases, spliceosomes, Hsp70 and Hsp60 chaperones, proteasomes, RNA degradasomes, protein translocases, reverse transcriptases, aminoacyl-tRNA synthetases, and error-correcting machines.

Many codes are mutually dependent in order to function, and cells could not have evolved each decoder sequentially. Ribosomes require the mRNAs from DNA-dependent RNA polymerases, but these polymerases are composed of protein products from ribosomes. Both decoders require the energetic ATP molecules from ATP synthases, which themselves cannot exist until ribosomes and RNA polymerases already work. Being coded information systems, additional guidance is provided in cells through engineered components such as the cytoskeleton, lipid rafts, membranes, pores, chemical gradients, correct placements of synapses, correct binding strengths, and nuclear subcompartments.

Computer architectures structure long-term storage capacity hierarchically to process data at different levels of granularity: data centers and distributed file systems; hard discs; disk partitions; files and extents; tracks; sectors and data blocks; and bits. In cells the same kinds of hardware principles are observed: ecologies of cooperating bacteria and multiple cells in eukaryote organisms; genomes; chromosomes, plasmids, mitochondria, and chloroplasts; euchromatin/heterochromatin and DNA looping; DNA regions defining primary RNA transcripts; exons/introns; and nucleotides.

Cells must be interpreted as holistic systems whose origin cannot be explained by neo-Darwinian theory.

Introduction

In Part 1 (Truman, 2016) of this two-part series, we showed that formal data structures, generic programming constructs, and file formatting are used in both computer programming and cells to process logic. Many independent codes exist in cells, each developed from elementary symbols and used by their own programs. These codes rely on variables and values and, reminiscent of computers, have well-defined data types and a list of valid operations that can be performed

* Dr. Royal Truman. Mannheim, Germany, royaltruman@yahoo.com
Accepted for publication June 20, 2016

with them. In both cells and computers, variables or fields can be used to create indices and concatenated indices (for example cis-regulatory elements) to identify the location of data (the portion of DNA to transcribe).

Data values can be assigned to variables (such as when a receptor is suitably bound by a ligand) and stored in linked lists (as seen in data processing by RNA and DNA polymerases) and arrays (for example translation of mRNA codons). The location of data also can be identified by "pointers" to memory addresses (as occurs in cellular signaling cascades). The various programs consist of formal software elements such as Boolean logic; iteration with clearly specified start and termination rules; control structures such as "GoTo," "wait," and processing commands permitted for that data type. Computers and cells also implement short- and long-term data storage, parallel processing, source code sharing and reuse, data compression, and data archival.

These software elements permit intended outcomes to be specified using formal symbolic logic. Abstract variables can be manipulated; for example: if x=true and y=2 then {do the following}. The stepwise instructions can be defined without considering the physical implementation.

Here in Part 2 we now consider how the cellular programs get instantiated, showing once more that there are remarkable parallels with how data is physically organized and used by computers. We will emphasize the need for very complex cellular hardware to process the different codes and how these are part of the holistic design of adaptable, self-correcting, and self-replicating cells.

Distribution and Partitioning of Available Storage Capacity

Data Storage Formatted Hierarchically

Significant advantages result from creating architectures to structure how data is stored physically. These structures are transparent to the processing logic but can affect the speed and reliability of execution. Computers and cells use well-defined standards to store data in a hierarchical manner, with each level including one or more members of the subordinate level (Figure 1). Although the analogies suggested in Figure 1 display remarkable similarities between computer and cellular technologies, they should be viewed as illustrative only. Computer technologies could change in the future, and not all cell types are identical (e.g., prokaryotes/eukaryotes). In the computer-versus-cellular analogies, some of the layers operate independently. The location of database data blocks is independent of physical sectors, and parts of exons could be located within a single nucleosome (and vice versa). In all cases, there are good design reasons for organizing parts of the data in specific manners, and special read/write technology is necessary for all layers of the architecture.

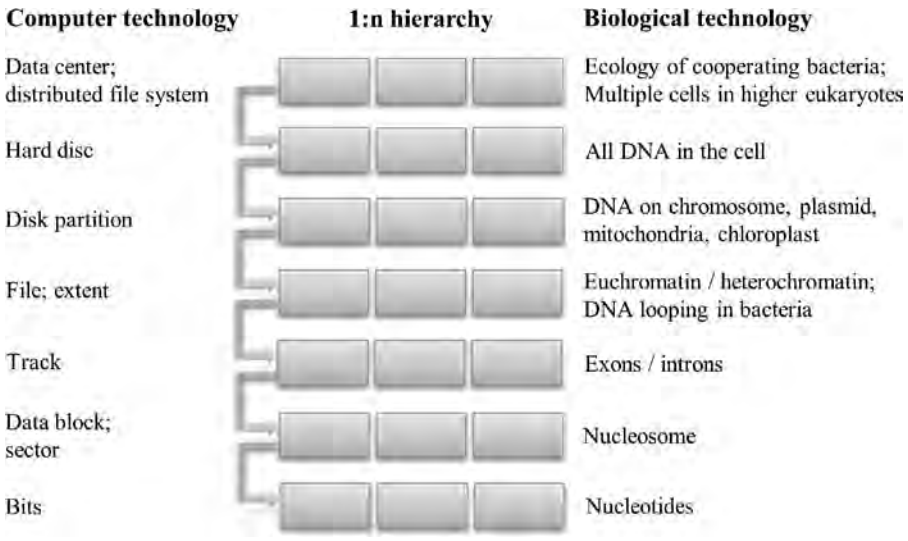

Figure 1. Computers and cells organize storage capacity hierarchically, with each level containing multiple members of the embedded sublevel.

Data centers and distributed file system

Different and multiple copies of computer data are often stored separately, on disc arrays, separate servers, and data centers. In an analogous manner, ecologies of bacteria store both different data and multiple copies of the same genome. Higher eukaryotes also store different data in separate organisms and multiple copies in different cells within the same organism. By cellular data we are not limiting ourselves to DNA. In Part 1 (Truman, 2016) of this series, we pointed out that coded variables and values are stored on DNA, RNA, proteins, sugars, membranes, and other biochemical substances.

Hard disc

It often makes sense to organize essential data in one easily accessible location. Storage media such as hard discs are used by computers linked to hardware processors and external devices. In the case of cells, such data are also placed together within the organism and linked to data processors, everything within a protective membrane to separate from the environment.

Disk partition

Long-term storage capacity for computers is often separated into partitions, which act as separate disk drives. Data can often be transferred between partitions, and sometimes they contain different operating systems or types of data. This is also true of cells, which organize the available DNA into separate chromosomes, plasmids, mitochondria, and chloroplasts. Partitioning must be unambiguous. This is implemented in plasmids by making them circular. Initiation of chromosome replication in prokaryotes is defined by sites called *origins of replications* and a special *terminus* regions identity where replication terminates. In eukaryotes, chromosome initiation sites called *autonomous replication sequences* (Shapiro, 2011, p. 36), along with centromeres (O'Sullivan and Karlseder, 2010; Sullivan et al., 2001) and telomeres (the caps at the ends of chromosomes) (Sadaie et al., 2003) identify individual chromosomes.

File, Extent

A disk partition can contain many files, or database extents. Similarly, chromosomes in eukaryotes are organized into smaller portions called euchromatin and heterochromatin (Shapiro and von Sternberg, 2005; Shapiro 2006). A computer file may contain records of varying lengths (e.g., in databases), even as sections of DNA euchromatin can contain multiple transcription regions. In bacteria, a similar principle is performed using DNA loops created through bound proteins, which lead to physical segmentation of the DNA (Cournac and Plumbridge, 2013).

Track

A track forms a circle on the hard drive platter and represents only a small portion of the entire data on the hard disk. In a comparable manner, introns and exons are also subdivisions of transcription regions.

Sector and data block

Sectors and database data blocks typically contain a fixed number of bits. A database extent can contain several blocks, and a track contains several data sectors. In eukaryotes nucleosomes also represent small, fixed-length data storage units within transcription regions.

Bits

The smallest entity used digital computers are bits, and for DNA this corresponds to nucleotides.

External devices

In addition to structuring data hierarchically, electronic data can be stored on external devices such as printers, scanners, and cameras. Cells can also interact with DNA stored externally, such as on viruses and microRNA's in vesicles travelling the blood system. Furthermore, DNA on bacteria associated with higher organisms provide them with many valuable functions.

Design Reasons for Structuring Storage

Process data at different levels of granularity

For different purposes, such as storage, retrieval, and processing, the data sometimes involves large amounts of undifferentiated data and other times specific structured data. If a hard disc is badly corrupted and the damaged part cannot be repaired, then the entire content must be replaced. But sometimes an identifiable subset only was read or written incorrectly—such as a partition or data block, or IP packet—and the correction should be executed at the correct level of granularity. In a similar manner, a severely damaged organism ("data center") may need to be replaced, or only a damaged single cell, or an alternative plasmid copy should be used, or only sections of mismatched DNA strands.

Concurrent access

Permitting concurrent access to different data removes processing bottlenecks, which requires that relevant subportions of the data be demarcated. Multiple RNA polymerases can transcribe simultaneously on the same or different chromosomes, different ribosomes can translate different (and sometimes the same) mRNA, and cells can work independently of each other.

Local and distributed processing

Sometimes processors need to be moved to where data can be collected, as in process controllers linked to manufacturing sensors. Analogously, bacteria with their data processing equipment need to be located near nutrients.

Other times data needs to be transferred to where it can be processed, as in computer client server architectures. In cells, various genes required for a specific function are often brought together for processing in transcription factories (Mitchell and Fraser, 2008; Melnik et al., 2011).

Process different kinds of data on different physical media

Structuring data correctly, including file-type formatting discussed in Part 1, permits specialized software to process raw data for images, sounds, spreadsheets, and so on effectively. In an analogous manner, different codes in cells require that DNA, RNA, protein, and other media be physically prepared properly to permit data to be stored and retrieved from them. In this paper, we limit detailed analysis to how DNA and RNA are organized to optimize logic processing by difference programs. Future work will reveal how other information carriers are structured, such as for proteins (motifs, domains, multi-domain complexes, etc.).

Effective interaction between long- and short-term memory and backup copies

Instead of loading terabytes of irrelevant content from a hard disk into short-term

memory, it makes more sense for an application such as a database query to read a limited number of relevant blocks into memory. It is also more efficient to read only once a large amount of data that includes a substructure of relevant parts. Recreating the transcription apparatus (ribosomes) for each individual protein needed is less efficient than to activate multiple relevant genes concurrently, as done by operons and transcription factories (Rieder et al., 2012).

Available memory in computers and cells can be reorganized and freed up for other purposes. RNA no longer needed is degraded to free up "short-term memory" and re-created when needed from only small portions of chromosomes.

Cellular Hardware Plays Multiple Roles

The software constructs discussed in Part 1 are used to express processing needs, independent of the hardware. But at some point, a physical implementation is necessary to be of any use. Computers require many hardware components to work, and not only processors, to interpret the coded data. Electrical power is delivered via cables to the correct location, temperature needs to be regulated, and sensitive parts protected from the environment.

Physical structures in cells must also provide many services, in addition to interpreting the codes. Hardware such as membranes, pores, and protein filaments can channel outcomes to produce the intended results, which is a feature of coded information systems (Truman 2012a, 2012b, 2012c 2012d, 2013, 2015). This guidance, like the effect from coded messages, decreases the entropies with respect to a reference state. Hardware components in cells, however, must do far more than in computers. Cellular hardware must be adaptable and dynamic with respect to when and where they are made available, how long they exist, and even the number of processor copies.

Hard and Software Independence in Digital Computers

Since the seminal Turing machine model created by British mathematician Alan Turing (1912–1954), hardware engineers can work largely independently from software programmers. In Turing machines, symbols are stored on a readable tape using a simple head reader and a small rule set. Modern digital computers are generally based on the more refined Von Neumann or the Harvard architecture, using stored programs with instructions, processed by separate hardware that performs the calculations (Bray, 2009, p. 40).

No human computer technology can match the sophistication of cells, which self-repair, replicate autonomously, and manufacture most of the components needed (including their own energy supply). Cellular decoders process the variables and values used by the different codes found on DNA, RNA, proteins, and sugar, as discussed in Part 1. But cells can do many things no computer system can match. They can regulate when, where, and how many processors (for example DNA polymerases, tRNAs and ribosomes) are concurrently in use. This can depend on a developmental schedule, be a response to current biological needs, and be affected by external signals.

The Turing (Turing, 1950), von Neumann, and Harvard computer models provide a partial picture for how cells process logic, being unable to take into account the direct physical participation of the genome in reproductive and regulatory interactions (Shapiro, 2013, p. 289). These models do not naturally capture the cellular reality of multiple codes, often overlapping the same physical location and requiring many kinds of integrated decoding processors. Nevertheless, at a fundamental level cellular codes are being processed as part of a complex program with high fidelity, and the insights gained from digital and analog computers are useful to understand why cells are not simple mechanical machines. We will distinguish next between cellular hardware which process coded parameters from hardware that do not use a code to guide their operations.

I. Hardware Processors Necessary to Decode

In Part I (Truman, 2016) we discussed how variables are restricted to data types, such as integer, alpha numeric, and Boolean for computers, and that many data types are used in cells, constructed from elementary symbols having definite electronic and geometric characteristics.

A key to understanding cells is to recognize that different codes are used to establish linked chains of sending and receiving variable values. Each code uses variables that can assume values. One region of a ribosome defines a variable that can assume one of 65 values (a codon or be empty). A different region of the same ribosome defines a variable able to assume the value of one of 20 amino acids on the growing chain. RNA nucleotides can also be used to define variables, such as in the intron/exon splicing and other codes.

Each cellular code requires its own kind of decoder. Ribosomes cannot work with splicing signals recognized by spliceosome, nor can spliceosomes work on the basis of codons.

The principle becomes apparent. Cis-regulatory element binding sites (CREs), protein phosphorylation sites, methylated DNA, hormone receptors, mRNA sites recognized by miRNAs, etc. each use their own distinct variables. These codes can only be processed using their own unique decoding hardware machinery, the variables possess a unique data type, and they are limited to a set of allowed operations.

Data Structures Require Precise Engineered Components

In Part I, we discussed the need for software logic to express intentions correctly, and we now discuss the decoding hardware also necessary for the informational system to work. The molecular machines involved must be able to do the right (and avoid doing the wrong) thing for all the codes used in cells. For example, in eukaryotes spliceosomes must identify precisely the nucleotide ends that define the intron/exon boundaries to create up to thousands of possible protein variants, as in the case of neurexin genes in nerve cells or the protein Dscam (Bray, 2009, p.158). Rampant false positives would be deadly.

Molecular machines require very low amounts of energy (Mavroidis et al., 2004) and lie at the heart of every significant biological process (Erbas-Cakmak et al., 2015; Alberts, 1998; Glogocka et al., 2015). Cells are not a bag of enzymes and biochemical reactants but regulated systems working on individual molecules (Głogocka et al., 2015 p. 248). This contrasts with chemistry as performed by humans:

> In stark contrast to biology, none of mankind's fantastic myriad of present-day technologies exploit controlled molecular-level motion in any way at all: every catalyst, every material, every polymer, every pharmaceutical, and every reagent all function through their static or equilibrium dynamic properties. (Erbas-Cakmak et al., 2015 p. 10157)

Cellular machines have, however, inspired the growing research fields of artificial molecular machines and nanotechnology (Balzani et al., 2000).

Let us consider the engineering challenge. For a future (evolving) spliceosome, virtually any nucleotide sequence pattern could be "chosen"—of practically any length—to define the intron/exon boundaries (and the recognition pattern need not be limited to contiguous nucleotides). Therefore, there are vastly more wrong candidates where a developing spliceosome could tear up DNA strands than a correct pattern that by astonishing good luck happens to delimit useful exons throughout the genome to produce valuable proteins (as individual or multiple exons). An average-size exon is only about 137 nucleotides long, surrounded by a sea of large introns in the pre-mRNA (Berget, 1995; Ward and Cooper, 2010). An indiscriminant evolving spliceosome would almost always produce only interfering mRNAs, at great cost of energy and material.

The same considerations apply to the other cellular decoders, such as DNA and RNA polymerases. Let us consider translation. Nature can have no foreknowledge how to process mRNA or how to develop the genetic code. Nucleotides are chemically complex, and every portion has the potential to interact physico-chemically with other substances.

Suppose that after countless evolutionary attempts a proto-ribosome evolved able to translate all the codons correctly with probability of 0.5 (or incorrectly but with neutral effects) an incredible feat considering there are 20 different amino acids. This would be some point during a Darwinian process of optimization. Assuming a small protein size of 200 residues, binomial probability calculations done with Microsoft Excel reveal that less than two mRNAs out of a million ($p=1.8x10^{-6}$) would have 2/3 or more of the 200 codons (i.e., 133/200 or more amino acids) translated correctly or acceptably (Figure 2). The cell would not be viable even using such optimistic assumptions. A minimum cell requires many, many copies of hundreds of different kinds of functional proteins and no cell lineage could survive such chaos.

We conclude that from the very beginning any ribosome must display astonishing fidelity. The implausibility of evolving hopelessly deficient codes can be applied to many other examples, such as to specify what part of DNA should be transcribed to produce mRNA and when; and successfully distributing only one member of each chromosomes to daughter cells.

Examples of Some Processors

Having concluded decoders must be very precise *ab initio* to offer any benefit, we will now draw attention to how complex these molecular machines are, using some examples.

Spliceosomes

The spliceosome ribonucleoprotein machine consists of about 300 distinct proteins and five RNAs (Nilsen, 2003; Chen and Moore, 2014; Matera and Wang, 2014; Ward and Cooper, 2010). It is believed the spliceosomes are constructed de novo for each round of splicing in an ordered but immensely complex stepwise assembly (Lodish et al., 2000, pp. 413–426). In their review Chen and Moore (2014) refer to "the incredible complexities facing structural biologists bold enough to even contemplate complete structural understanding of the splicing machinery within their lifetimes" (p. 147).

Signals that identify introns on the primary RNA include short sequence motifs at the 5' and 3' splice sites (typically a GU and AG respectively); a branch point sequence that contains an adenosine; and the polypyrimidine tract (Figure 3). A typical mammalian gene has 7–8 exons spread out over ca. 16 kb. The exons are relatively short (~100–200 bp), and the introns >1 kb. A typical processed mRNA consists of ca. 2.2 exons. Trans-splicing occurs when exons from different transcripts are linked, encoding chimeric proteins (Lasda and Blumenthal, 2011; Michaeli, 2011; Frenkel-Morgenstern et al., 2012; Herai and Yamagishi, 2010).

Ribosomes

Ribosomes, where translation of mRNAs occur, are composed of 54 ribosomal

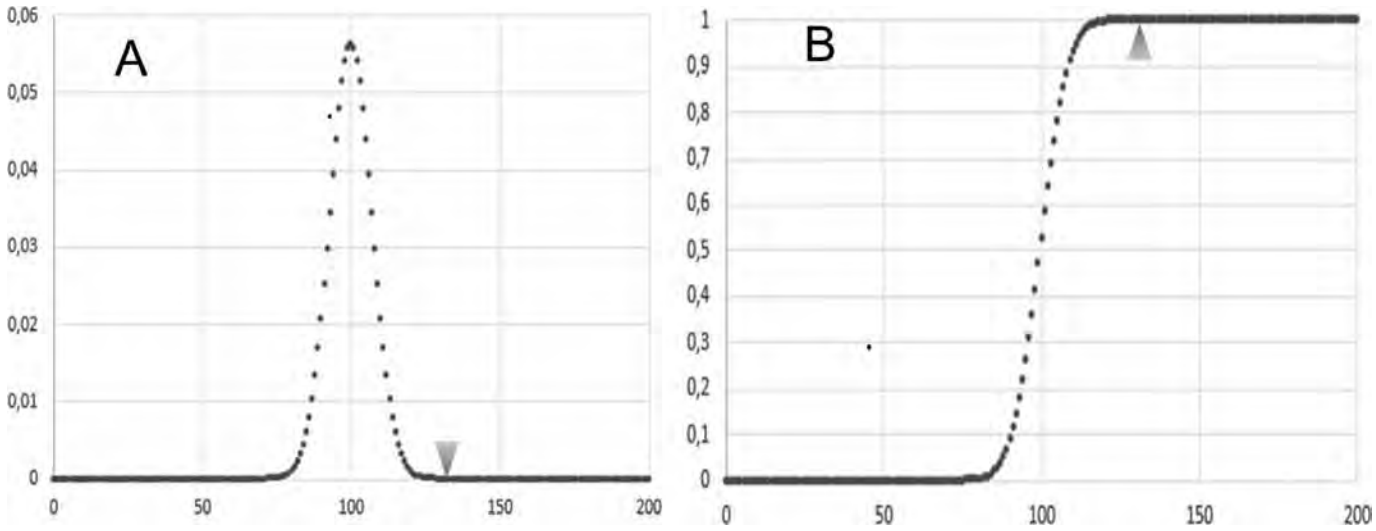

Figure 2. Probabilities for 200 codons to be translated correctly (or wrong but with neutral effects), assuming each codon has an average individual p=0.5 of success. From binomial probability statistics, this leads to mean = 100, σ =7.07. Arrows show the location of 133 successes.

A: Probability of 0, 1… to 200 successes for the 200 codon attempts.

B: Cumulative probability of success for the 200 codon attempts. For the 200 codons, 52.8% will have between 0 and 100 translated correctly, and 99.9999% will have $^2/_3$ or less.

During an earlier evolutionary period having an average probability p=0.1, over 99.9999999% of a 200-codon mRNAs would have only 0 to 50 codons translated correctly (or wrong but with neutral effects), i.e., one-fourth of the mRNA's codons (mean = 18, σ=4.24).

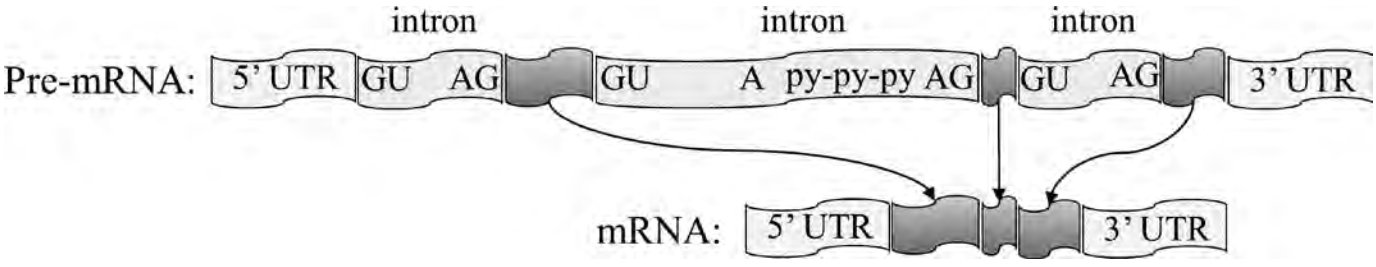

Figure 3. Results after a spliceosome has removed introns. Introns are usually delimited by GU and AG at each end and additional nucleotide patterns near the intron/exon boundary help identify the key di-nucleotides. A key internal A and polypyrimidine tract are also involved.

proteins and 3 large RNA molecules in prokaryotes (Chen and Williamson, 2013; Staley and Woolford, 2009) and 79–80 proteins and four ribosomal RNAs (rRNAs, >5500 nucleotides) in eukaryotes (Wilson and Doudna Cate, 2012).

Ribosomes are worthless until all the multidomain aminoacyl tRNA synthetases (aaRS) are also fully operational in addition to the proteins needed to create these enzymes. We observe that charging the tRNAs requires another, nongenetic code (Trifonov, 2011; Hou, 1988). Twenty different types of aminoacyl tRNA *synthetases* (aaRSs) are found in all organisms, each responsible to covalently link the correct amino acid to the correct tRNA. These aaRS differ in length, three-dimensional structure, molecular weight, and subunit organization, and are very different in amino acid sequence (Panwar and Raghava, 2010).

For translation to work well enough to have any value, a large number of other enzymes are also needed for quality control purposes. These are very precise machines that ensure error rates between 1/2400 to 1/40,000 although the differences between amino acids are miniscule, such as a single methylene group (Nureki et al., 1998; Sarfati, 1999).

Proteasomes

Proteasomes degrade unneeded, damaged, and misfolded proteins in eukaryote cells and some archaebacteria. In eukaryotes, the proteins to degrade are identified by conjugation to ubiquitin proteins (Amm et al., 2014). The eukaryote 26S proteasome consists of 33 distinct subunits (Besche et al., 2009), including the 20S proteolytic core chamber (where protein degradation occurs), which is attached to two 19S regulator caps on each end. These caps consist of about 20 different proteins, which serve to capture and guide proteins into the chamber. 19S caps and ubiquitination have not been found in archaebacteria.

Proteasomes can also interact with the immune system by generating short peptide fragments from pathogens that are used as antigens on the surface of lymphocytes to mount an immune response (Wang and Maldonado, 2006).

DNA Polymerases

DNA polymerases are involved in several key processes, such as DNA replication, many kinds of DNA repairs (Ball et al., 2014), translesion DNA synthesis, DNA recombination, cell cycle control, DNA damage checkpoint functions, and chromatin remodeling (Hübscher et al., 2000; Wu et al., 2014). DNA polymerases are highly accurate, with an intrinsic error rate of less than one mistake for every 10^7 nucleotides added (McCulloch and Kunkel, 2008; Kunkel and Bebenek, 2000).

Most DNA polymerases have a similar three-dimensional fold despite having virtually no sequence similarity (Federley and Romano, 2010). Five kinds of DNA polymerases are known in E. coli, 8 in Saccharomyces cerevisiae, at least 15 in humans, and 12 in Arabidopsis thaliana (Garcia-Diaz and Bebenek, 2007, p. 105), but even DNA polymerases in the same family often display very different properties (Garcia-Diaz and Bebenek, 2007, p. 105).

Some polymerases consist of a single polypeptide chain, and others, such as those involved in chromosome replication, are composed of several subunits. Even the single-chain versions are large and complex. For example, the first DNA polymerase discovered was Pol I, which is ubiquitous in prokaryotes. The E. coli version of the enzyme is composed of 928 amino acids (Brown et al., 1982). The processes carried out by polymerases are very demanding. For replication of the two DNA strands, the leading strand is synthesized in the 5' to 3' direction and the other lagging strand in the opposite direction.

Many more proteins are needed that collaborate with these large polymerases. For replication alone this includes those forming the origin recognition complex (ORC) so the polymerase can access the DNA, the replication protein A (RPA proteins), the MCM complex to unwind DNA, proteins to guide polymerase switching (on the lagging strand polymerase switching is necessary about every 100–200 nucleotides synthesized) and proteins to prevent premature dissociation from the DNA template (Cotterill and Kearsey, 2014).

RNA Polymerases (RNAP)

RNAPs are also very complex. Even bacterial RNA polymerases are very large. The core enzyme requires several proteins and consists of 5 subunits besides a transcription initiation factor sigma (σ) (Ebright, 2000). Using about 110 Daltons per amino acid and a median protein size of 267 for bacteria, the ~400 kDa core enzyme is equivalent to about 14 different proteins working as a unit. Figure 4 shows a strand of RNA being produced. Furthermore, in E. coli more than 100 transcription factors have been identified that modify the activity of RNAP (Ishihama, 2000). Multiple types of RNAP are found in eukaryotes (RNA polymerase I–V) (Ebright, 2000). RNA and DNA polymerases are so different that postulating an evolutionary common origin makes no sense.

Unique feature of cellular decoders

A series of papers presented at the 2012 Conference on Engineering and Metaphysics drew attention to the requirements for self-replicators to work (Mignea, 2014). The few examples of complex decoders presented above still fall short of expressing how sophisticated cellular technology is. In computers, the hardware is manufactured once and then reused multiple times to run programs until wear and tear renders it unusable. In cells, hardware processors such as RNA polymerases, spliceosomes, and ribosomes are built, disassembled, and then rebuilt over and over again

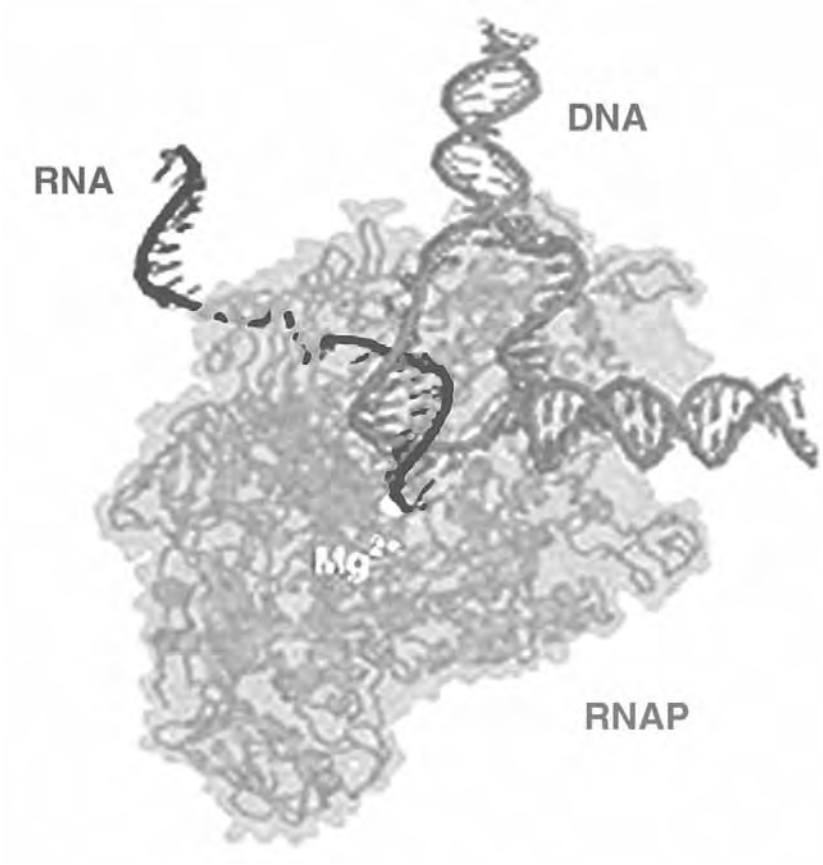

Figure 4. RNA polymerase (RNAP) during elongation (from Google Creative Commons, labeled "for commercial reuse with modification").

and at the location needed. This must be quantitatively regulated according to need and cell type. There are approximately 20 thousand ribosomes in E. coli, and up to 10 million in human liver cells (Woodward and Gills, 2012, p. 43). This is a general observation for cells, which regulate when, how many, and where the decoding processors are built, reliably and over thousands of generations.

Optimization trade-offs are at play. Reliably repairing entire molecular machines damaged by bond breakage, oxidation, and other chemical reactions is probably not feasible as a design concept, so rebuilding from high-quality components makes sense. There are also cases where the biological information processing equipment, once formed, remains intact during the remainder of the organism's lifetime to constantly process signals, like the whisker neurons of mice, which link through a complex relay path to the cerebral cortex (Kirschner and Gerhart, 2005, p. 161). This also makes sense, since whiskers are important sense organs needed continually to navigate in the dark and tens of thousands of perfect copies are not needed.

Classification of Molecular Machines

The cellular hardware that process the variables and their values in the different codes are examples of molecular machines (MMs). New MMs are being reported continually (Alberts, 1998; Luskin, 2010), and there are hundreds of MM types. We wished to find some way to classify MMs to help understand cells as a whole. An evolutionary origin for cells would require this, since the starting point would have to be as simple as possible. Are some logic processors so indispensable they must have arisen earlier, and could others be added later?

We tried to develop a scheme based on necessary precedence, observing that some MMs require a preexisting foundation of other MMs. F_0F_1 ATP synthases can be built only if the necessary proteins are already available, suggesting ribosomes are more fundamental. However, we know that translation at ribosomes requires ATP to function, which implies the opposite precedence.

Additional examples we considered were also inconsistent with a hierarchical classification. Ribosomes function only once the whole complement of aminoacyl-tRNA synthetases (aaRS) are in place and mRNAs are being produced. However, both aaRS enzymes and RNA polymerases are constructed from proteins, so functioning ribosomes would need to already have existed.

This is significant since it implies an all-or-nothing design. This is inconsistent with an unguided evolutionary scenario, whereby simpler foundational MMs could somehow evolve first to prepare the ground upon which more complex MMs can arise. These foundational MMs, though, require the "higher-level" MMs. We must not overlook that foundational MMs such as aaRS offer no biological value alone. Attaching activated amino acids to tRNAs without a genetic code or ribosomes serves no purpose, and any organism wasting the raw materials and energy thus consumed would be at a selective disadvantage.

We propose tentatively a four-category classification of cellular information processors, where at least some members within each category can interactively

affect others within the same or the other categories (Figure 5).

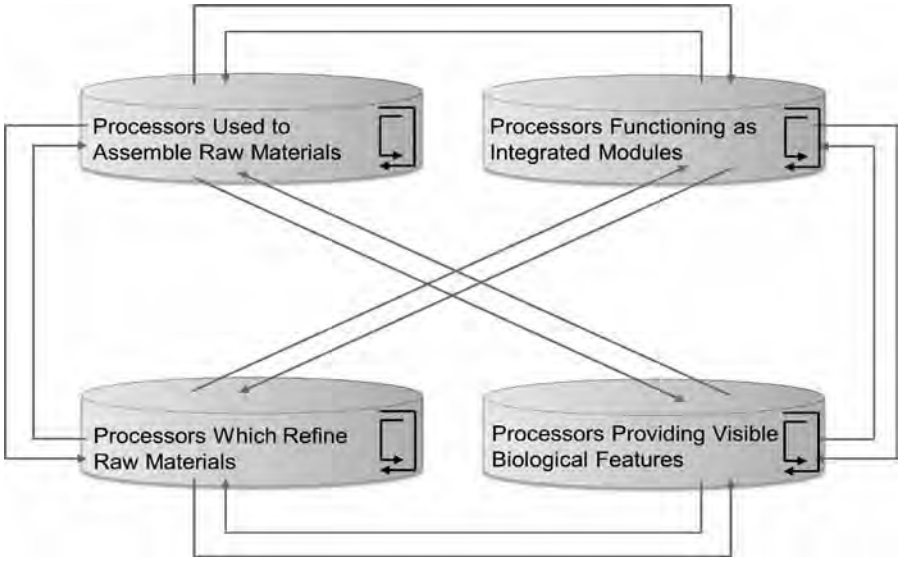

Figure 5. Model to classify cellular information processors. Members from each category can affect and be affected by the results of processors from the other three processors.

i) Processors Used to Assemble Key Raw Materials

MMs in this category assemble components into more complex building blocks (e.g., proteins, DNA, RNA, lipid, and sugar polymers) and the MMs that manufacture high-energy molecules such as ATP, Pcr (Phosphocreatine), NADH, FADH2 and NADPH. Examples include **ribosomes** (Staley and Woolford Jr., 2009); **DNA polymerases** (Garcia-Diaz and Bebenek, 2007; Federley and Romano, 2010; Hübscher et al., 2000; Wu et al., 2014); **DNA-dependent RNA polymerases** (Jonkers and Lis, 2015); **RNA-dependent RNA polymerases** (Ahlquist, 2002; Maida and Masutomi, 2011); **reverse transcriptases** (Ducrest et al., 2002); and aminoacyl-tRNA synthetases (**aaRS**) to charge tRNAs (Yadavalli and Ibba, 2012). We provide only some examples here and will not go into details but reiterate how many proteins are typically involved per MM. For instance, the aaRS core complex in multicellular eukaryotes is composed of 8 proteins with 9 enzyme activities and 3 auxiliary proteins (Norcum et al., 2005).

ii) Processors That Refine Raw Materials

Several molecular machines refine the location, concentration, half-lives, and other properties of biomolecules already assembled. Examples include **spliceosomes** (Nilsen, 2003); **hexameric AAA+ enzymes** (Hanson and Whiteheart, 2005; Bar-Nun and Glickman, 2012); **RNA helicases** (Jankowsky, 2011); **DNA helicases** (Sommers et al., 2015); **topoisomerases** (Nitiss, 2009); machines producing chromosome double-strand breaks followed by **homologous recombination** during meiosis (Baudat et al., 2013; Brick et al., 2012; Yamada et al., 2013); protein kinase c (**PKC**) (Steinberg, 2008); **proteosomes** (Bhattacharyya et al., 2014); **apoptosomes/caspases** (Rastogi and Sinha, 2009); **RNA degradasomes** (Bandyra et al., 2013) including RITS complexes (**R**NA **I**nterference by **T**ranscriptional **S**ilencing) (Verdel et al., 2004; Verdel et al., 2009; Sugiyama et al., 2005); **kinesin motors** (Marx et al., 2009); **myosin motors** (Sweeny and Houdusse, 2010); **protein translocases** such as Tim/Tom Systems (Schmidt et al., 2010; Koehler et al., 1999); **ClpX** (a molecular machine that unfolds proteins and then transports them into another complex) (Baker and Sauer, 2012); the bacterial **SecYEG PreProtein** translocation channel (Lycklama and Driessen, 2012); **Hsp70** chaperone machines (Mayer and Bukau, 2005; Mayer, 2013); and **Hsp60** chaperone machines (Fink, 1999; Merendino et al., 2010).

iii) Processors Providing Services as Integrated Modules

Some MMs provide holistic cellular services that go beyond assembling and modifying biochemicals. Examples include **calcium pumps** (Brini et al., 2013); **cytochrome C oxidase protein pumps** (Wikström, 2004; Yoshikawa et al., 2011; Yoshikawa and Shimada, 2015); **meiotic and mitotic cohesin complexes** to ensure accurate chromosome segregation during replication (Mehta et al., 2012; Maddox and Ladouceur, 2015); **condensin I and II complexes** (Hirano, 2012); **immunological synapses**, a complex cellular structure that forms at the interface of a T cell and a cell that expresses the peptide–MHC complexes (Dustin, 2014); **Smc5/Smc6 complexes** (Outwin et al., 2009; Farmer et al., 2011); **cytoplasmic dynein motors** (Allan, 2011); **MRX complexes**, which determine telomere integrity by a protein-counting mechanism (Lue, 2009; Ball et al., 2014); **F_0F_1 ATP synthases** (Long et al., 2015), which consists of at least 22 subunits (Nakamoto et al., 2008); bacteriorhodopsin (**bR**), the simplest known light-driven proton pump (Haupts et al., 1999); **hemoglobins** (Schechter, 2008); and **kinetochores,** for

```
public class TranslateCodons {
  public void ProcessEachCodon(String codon){
  String c = codon;
  switch (c) {
    case "GCU": {
      // 1: Delay translation by time reflected by this codon. E.g. try {Thread.sleep(10);} catch (Exception e) {}
      // 2: Transfer Alanine from the tRNA to the growing chain
      // 3: Release the attached tRNA from the codon
      // 4: Move the mRNA forward one codon position
    }
    case "CGU": {
      // 1: Delay translation by time reflected by this codon.
      // Java example try {Thread.sleep(35);} catch (Exception e) {}
      // 2: Transfer Arginine from the tRNA to the growing chain
      // 3: Release the attached tRNA from the codon
      // 4: Move the mRNA forward one codon position
    }
    //case n Code the rest of the cases
    }
  }
}
```

Figure 6. Some of the instructions being carried at the *A* site of ribosomes, identified by empirically observing what happens when each codon on mRNA is processed.

which over 80 protein components are involved (Cheeseman and Desai, 2008). These machines need not be physically connected into adjacent, cohesive units.

A matter to consider is whether there is a clear boundary between multicomponent MMs and integrated processes such as the Calvin-Benson cycle (11 enzymes; Michelet et al., 2013) and the TCA cycle (King, 2015).

iv) Processors Providing Visible Biological Features

Finally, it seems that some MMs comprise biological features whose value and operation are clear by simple observation even lacking understanding about the molecular mechanisms involved. Examples include **bacterial flagella** (Xie et al., 2011); **eukaryotic flagella and cilia**, using over 360 different proteins in the case of green alga Chlamydomonas reinhardtii, (Pazour et al., 2005); **blood clotting** cascades (Adams and Bird, 2009); **antibodies** and the adaptive immune system (Iwasaki and Medzhitov, 2015; Schenten and Medzhitov, 2011; **apicomplexan glideosomes** (Boucher and Bosch, 2015; Frénal et al., 2010); type I-VII chaperone–usher and the curli **secretion systems** (Costa et al., 2015; Douzi et al., 2012; Chatterjee et al., 2013).

Final Comments on Type I Cellular Hardware

The type I hardware discussed above is indispensable to process coded data in a cell that "either instructs or directly produces nontrivial function at its destination" (Abel, 2011, p. 3). In Part 1 (Truman, 2016) we mentioned that in the genetic code, different codons can specify translation to the same amino acid, but these codons are not 100% interchangeable for all biological purposes. There are at least five "variables" known on ribosomes where logic processing occurs: at the **A** (aminoacyl), **P** (peptidyl), **E** (exit), **T** (elongation factor Tu), and **I** sites (initiation). At the **A** site the value of the variable is specified each time by a codon from an mRNA placed there, and then several logical operations are performed (Figure 6). One operation communicates which amino acid to add to the growing peptide chain.

RNA transcription elongation rates are also highly variable throughout different genes, which affect co-tran-

scriptional processes such as splicing, termination, and genome stability. This pausing, especially for Pol II, involves several transcription factors and complexes such as the NELF and DSIF in gene and sequence-specific ways (Jonkers and Lis, 2015). RNA interference is another example of an RNA code specifying processing steps. Binding of miRNA to excessive mRNA can prevent translation but may also recruit enzymes to actively degrade the mRNA (Francis, 2011, p. 130).

We now observe something remarkable from a computer design perspective. "Instructions" in cells such as how fast to translate each codon, to release the attached tRNA, and to move the mRNA forward one position, are not implemented using prescriptive, readable instructions in a source-code program but through precisely crafted biochemical features. These include the bonding strength between the tRNA's anti-codon with the mRNA codon, the concentration of that particular tRNA present, and attachment of a high energy ATP at just the right time and place.

II. Hardware that Contribute without Codes

Two kinds of computer and cellular hardware are used to guide outcomes: those whose purpose is to process coded data (as discussed above) and those that operate primarily without codes, which we now consider.

In cells, far more hardware functions are integrated into the entire system than in computers. It is the nature of computers to provide value by running software programs that generate conclusions in symbolic form that humans or other programs to a major extent then act on. Individual self-replicating computers do not exist, needing to synthesize their own energy and their own internal hardware (cables, memory cards). Cells are to a much greater extent "their own customers" and need to physically act on the symbolic conclusions made. It is not enough to calculate how much of which materials or energy need to be generated, or that repairs must be performed, or the timing of cell cycle phases. The cells must physically act on most of the consequences from these conclusions on their own.

Insights from CIS Theory

Human DNA consists of four possible nucleotides at around 3 billion base pair positions and bacteria of a few million base pairs. The coding capacity in terms of binary decisions is insufficient as the only source of guidance for organism development, continual decision making and internal cellular physical implementation, a concern that led to coded information theory (Truman, 2012a; Truman, 2012b; Truman, 2012c; Truman, 2013). Others are also concluding there are informational sources beyond what could be hard-coded on DNA.

Woodward and Gills (2012) wrote recently, "To be more specific, the zygote's three-dimensional structure seems supremely important. Every molecule, every structure, every atomic nook and cranny of the zygote potentially contributes to the cell's destiny" (p. 78). Francis (2011) also comments, "In the traditional view, genes function as executives that direct the course of our development. In the alternative view, which I advocate, the executive function resides at the cellular level and genes function more like cellular resources" (p. xiii).

DNA cannot directly specify all the things that occur such as the precise arrangement of all neurons, where and how which proteins interact, and the organization of the circulatory system, nor can it anticipate how to adjust our bodies for every step we might take.

Physical constraints designed into the CIS can decrease the contribution that needs to be communicated by coded messages by channeling possible outcomes. Here is an example: A coded message could instruct a robot which decisions to take to transverse a maze. As a simplified example, we are to provide a coded message instructing whether to move Up, Left, or Right (Truman, 2015).

In Figure 7 the eleven-symbol message "UUULUULUULU" is used, where every position communicates a new value to a receiver decision variable. Assume that upon entering any of the cubicles in the last row a sensor at that location is activated, attaching a rope to the robot, which is then dragged to the destination. The minimum number of *U*, *L* and *R* symbols required to reach the most distant square (labelled X in Figure 7) would be twelve.

The sensor and guiding rope could be cleverly placed closer to the origin, requiring less informational contribution from a coded instruction (Figure 8).

In the extreme case, the intention could be satisfied using the rope only (Figure 9). The rope could mimic the turns expressed formerly by coded messages.

What might be overlooked in this example, but is fundamental to understand cells, is that there is considerable design and specification agreement up front, and the software effort can benefit from this foreknowledge: the environment the code must operate in is set, how to interpret each symbol is known, and physical factors prevent undesired behavior from occurring (e.g., gravity prevents the robot from moving upwards).

Let us elaborate. We know the size of the squares and that they are the same, so each length of advance can be identical when designing the equipment and the code. This implies the precise duration to continue in each of the three possible directions does not need to be communicated along with each symbol. Knowing the nature of the now constrained problem and equipment design, no search algorithms need to be included, whereas this might not be true of other possible designs. If we know movement can only be in two dimensions, exactly 90° or straight ahead, then corrective

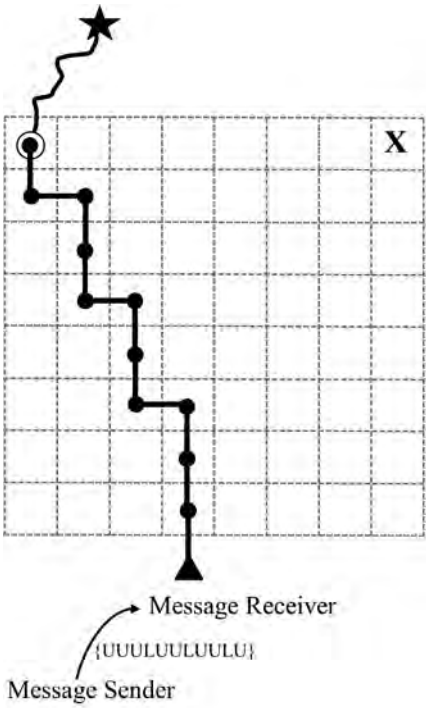

Figure 7. Coded messages consisting of *U*, *L* and *R* (Up, Left, Right) can instruct how to traverse a maze. At least twelve symbols would be needed to reach the most distant point labelled X.

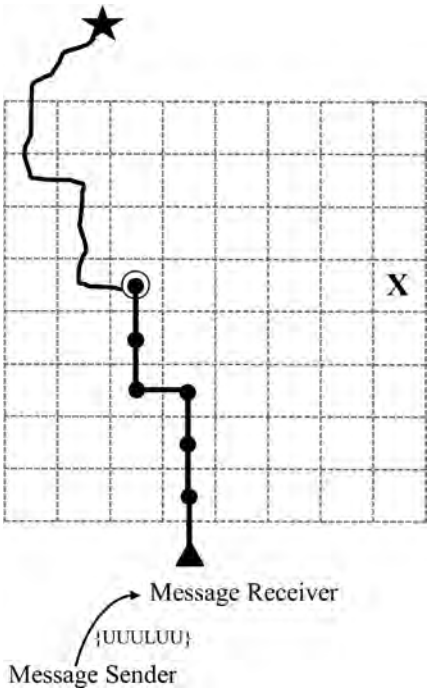

Figure 8. If the guiding rope is placed optimally in the fifth row, nine combinations of *U*, *L*, *R* suffice to reach the most distant point (marked X), from which correct exiting would be ensured with the help of an attached rope.

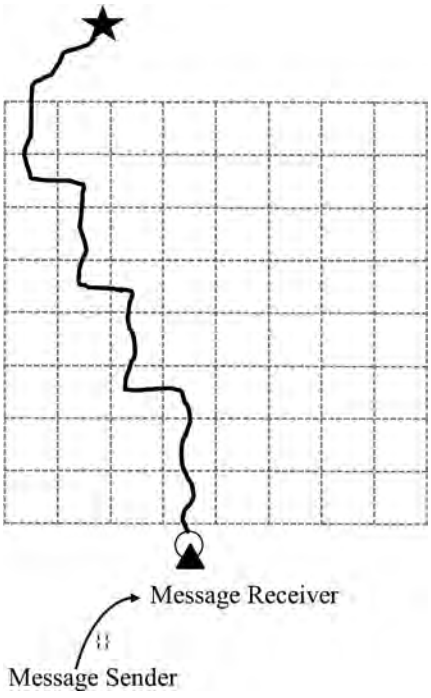

Figure 9. If the guiding rope is placed right at the entry point of the maze, no additional coded message would be necessary to traverse the maze.

instructions will not be needed. All these contributions permit sending the robot to any destination using very simple coded messages.

There are many biological analogies, such as when some mRNAs "automatically" delay translation or are stored and often transported to where translation should occur (Anderson and Kedersha, 2006). Another example involves ribosomes being created near where the mRNAs should be translated or are transported by kinesins along microtubule pathways. Other examples include the separation of chromosomes by spindles into precise locations for physical mechanical reasons before ending up in daughter cells during mitosis, and signal transmission along neurons. Movement in the maze example was constrained to a few adjacent squares in only two dimensions, but specifying in three-dimensional detail reliably where everything should be sent in a cell and when using only coded messages would be a nightmarish requirement.

Emergent properties do arise that need not be implicit in physical properties of individual components used by a system (Lehn, 2002). Turning the ignition key of a car initiates a series of processes that include combustion in a useful manner, resulting in "emergent properties." The same is true of biological systems, in which judicious preparation of the necessary components automatically results in the intended outcome without additional active coded informational guidance.

Although a multitude of coded parameters are used by cells to specify intention, they operate in the context of many designed physical constraints that significantly prevent erroneous behavior. Even diffusion of second messengers such as calcium, nitric oxide, carbon monoxide, lipophilic molecules (diacylglycerol, ceramide), superoxide, hydrogen peroxide, and hydrogen sulfide—presumably a Brownian type random movement—has been optimized in many ways. They are small and of appropriate solubility in a fluid environment to facilitate motion and will only need to explore restricted volumes using multiple copies and a high concentration of potential targets.

Cells possess several components that contribute toward correct (and

prevent incorrect) behavior. In the few examples mentioned next, we will see a recurring theme: the informative contribution of the assemblages is far greater than was provided by the coded messages used to put them together and to interact with them. There is a leverage effect. For example, a certain kind of membrane can usefully constrain a vast multitude of the appropriate kinds of bio-substances to a relevant region.

Cytoskeleton

The cytoskeleton is composed of proteins that can form long fibers and interact with cell membranes (Doherty and McMahon, 2008; https://en.wikipedia.org/wiki/Cytoskeleton). The designs in eukaryotic cells are particularly elaborate, where different proteins are used to form the microfilaments, microtubule, and intermediate filaments. About 70 proteins are used to build intermediate filaments, and all except crescentin are found only in animal cells (Herrmann et al., 2007). The cytoskeleton can interact with hundreds of other proteins, such as cross-linkers, capping proteins, nucleation promoting factors, and signaling molecules (Huber et al., 2013; Fletcher and Mullins, 2010).

The cytoskeleton provides indispensable services. Entire chromosomes are moved to daughter cells during replication (i.e., millions of individual nucleotides benefit as an ensemble), cell and organelle shape and volume are maintained, intracellular transport of entire organelles and vesicles are performed (i.e., with their entire biochemical content), the location of some macromolecules is constrained, portions of the cell are connected along which myosin-based motors can be guided, and the nuclear lamina is strengthened.

A few special signals guide the creation of these kinds of constraints. For example, microtubules grow in many directions until they encounter a signal on a membrane wall, at which point stabilizing agents lock the end of the microtubule in place (Kirschner and Gerhart, 2005, p. 151). As another example, vascular endothelial growth factors determine where new blood vessels are to develop, ensuring cells in most need of oxygen are supplied best (Kirschner and Gerhart, 2005, p. 169).

Pathways in retina to guide photons

In addition to the cytoskeleton discussed above, there are other examples of processes being guided by constraining physical pathways, such as the circulatory system, neural pathways of somatosensory perception, and brain neural networks. Photon signals that land on any of the estimated 260 million photoreceptors are transferred via 2 million ganglion cells on to the central nervous system (Gazzaniga et al., 2009, p. 178). Once the pathways are in place—during the organism's development and later learning—many of the activities that occur every second are largely automatic, and subsequent logic processing of the signals being received is thereby simplified. If photons were to land all over the place and not be guided to where interpretation is to occur in the brain, it would be close to impossible to make sense of these signals.

Organelles

As an example organelle, we will mention peroxisomes, which degrade fatty acids and eliminate toxic hydrogen peroxide in the cell. To do this, numerous enzymes are imported from the cytoplasm and concentrated together to work effectively (Montilla-Martinez et al., 2015). Here again we see the collaboration of CIS factors, since codes are also involved (Truman, 2016, Figure 1). Proteins to be transferred to peroxisomes carry certain signal sequences, called PTS1 and PTS2, which are recognized by import receptors. All the necessary participants are then forced conveniently close together within the organelle space.

Protein containing subcompartments

Ribonucleoproteins form discrete cytoplasmic RNA granules in germ cells ("polar" and "germinal granules"), somatic cells ("stress granules" and "processing or P bodies") and neurons ("neuronal granules"). These are involved in many processes, including mRNA localization and degradation, posttranscriptional modulation, and epigenetic gene expression (Anderson and Kedersha, 2009; Anderson and Kedersha, 2006; Kiebler and Bassell, 2006).

The eukaryotic nucleus contains subnuclear compartments such as the nucleolus, interchromatin granules (ICGs), PcG bodies, cajal bodies, perinucleolar compartment, and PML bodies, where special processes are coordinated (Spector, 2001; Shapiro 2011, p. 38; Osborne et al., 2004). Many steps in the DNA-damage repair process demand a careful spatial and temporal synchronization of events (Misteli and Soutoglou, 2009) carried out in "repair centers" (Lisby and Rothstein, 2005). The structures in the eukaryote nucleus, where specific and repetitive processing such as replication, transcription, and repair occurs, are referred to as "factories" (Mitchell and Fraser, 2008; Razin et al., 2011).

As with virtually all cell process, additional guidance within the subcompartments is provided by coded messages usually in the form of sequence patterns. There are signals called S/MARs (**s**urface/**m**atrix **a**ttachment **r**egions) that communicate the sites for localization to the nuclear lamina (Meuleman et al., 2013). However, it is also true that location and concentration of the correct components are ensured through purely physical principles once the engineered parts are assembled into a functional unit.

Lipid rafts

These are heterogeneous domains with high levels of cholesterol and sphingolipids on membranes that result in

different protein composition. The rafts are defined as "small (10–200 nm), heterogeneous, highly dynamic, sterol- and sphingolipid-enriched domains that compartmentalize cellular processes. Small rafts can sometimes be stabilized to form larger platforms through protein-protein and protein-lipid interactions" (Pike, 2009).

Lipid rafts are more ordered and tightly packed than the surrounding bilayer and seem to float in the membrane bilayer. Although more common in plasma membrane, lipid rafts have also been identified in lysosomes and the Golgi apparatus.

This special environment within membranes can facilitate various processes in eukaryotes, such as cell signaling (by increasing the chances of the necessary protein-protein interactions), protein trafficking, and regulation of neurotransmission (Simons and Ehehalt, 2002).

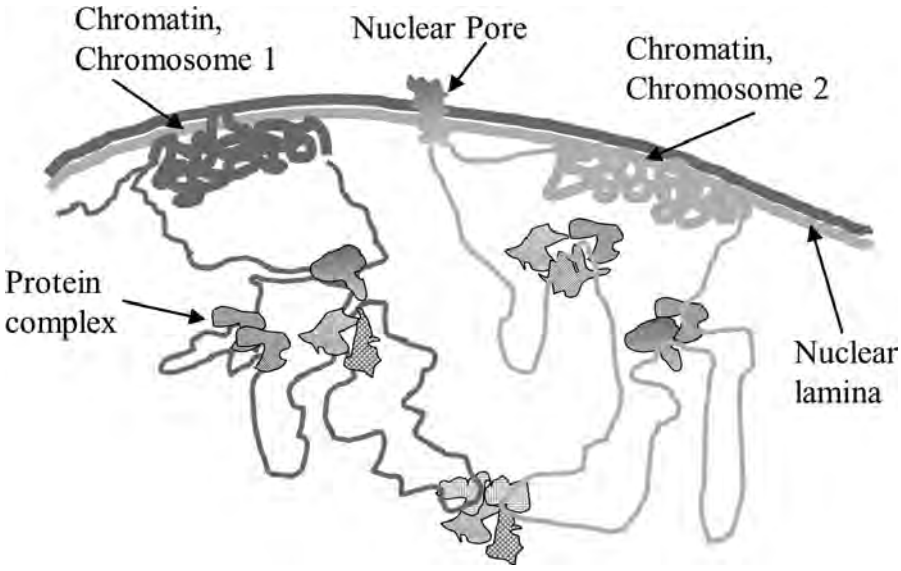

Figure 10. The surface of the nuclear lamina can help deactivate transcriptionally inactive chromatin (shown in heavier lines for two different chromosomes). Regulatory and active genes from the same or different chromosomes can be concentrated together in the nuclear interior using protein complexes.

Membranes and pores

Larger molecules such as amino acids, glucose, and nucleotides have a difficult time passing through membranes, so their location can be controlled by these barriers, alleviating the need for coded messages to specify this restriction in location. The surfaces of membranes can also be used for other purposes, such as helping to silence inactive chromatin (Walhout et al., 2013, p. 138) (Figure 10).

Eukaryotes have about 2000 nuclear pores—each composed of over 30 machinelike parts (Woodward and Gills, 2012, p. 40)—that permit transport of various substances in and out. In addition, ion channels permit the movement of ions like Na^+, K^+ and Ca^{2+} across membranes at a controlled rate, for example, by using a limited number and location of these channels.

Concentration gradients

Concentration gradients can act as informational signals. In the nervous system, nerve cells develop in many directions but soon undergo suicide (apoptosis) unless they enter a region possessing enough of a "survival factor" such as growth factors and cytokines (Fuchs and Steller, 2011; Portt et al., 2011).

Morphogens such as activin, bone morphogenetic protein (Bmp) 4, sonic hedgehog (Shh), and bicoid (Bcd) act as graded positional cues that control cell specification in developing tissues. Most are protein ligands that bind to transmembrane receptors and initiate intracellular signal transduction cascades to regulate the transcription of specific target genes (Ashe and Briscoe, 2006). To be a morphogen a concentration gradient of the substance must be able to generate at least two distinct cell types. These ligands can establish the initial polarities of embryos and specify cell identity in tissues by communicating their position with respect to the source of the signal.

Other concentration gradients produce their effects through more obvious physical-chemical reasons. Water moves automatically toward hypertonic regions (regions with high concentration of dissolved substances), and ions like Na^+, K^+ and Ca^{2+} naturally flow down their concentration gradients. (The Na^+/K^+ ion pump moves Na^+ out of the cell and K^+ in, against their concentration gradient, using energy provided by ATP.) Some passive transport is also facilitated by transport proteins in the membrane, but much guidance seems to be provided by the concentration gradient, a strictly physical factor that still needs to be properly organized.

Chemical binding

The need for coded instructions is decreased dramatically by ensuring that only a subset of possible chemical reactions can occur. Only the 3'5' phosphodiester bonds are permitted in nucleic acids to form the DNA and RNA backbone and only peptide bonds in the case of protein chains. Since the other reactions have deliberately been excluded from the cell design, the particular kind of reaction to perform

does not need to be specified by coded instructions.

Binding strengths

It is easy to overlook that cellular codes benefit by being able to make assumptions (recall the discussion above on the labyrinth). If the decoding procedures should be unreliable for any of the codes, the coded instructions would need to be far more complex, having to check current states and take corrective action. Current states are indeed checked by cells for complex processes, for example in the cell cycle. To illustrate, suppose the genetic code was highly unreliable for any of many possible reasons, one example being that ribosomes had not been precisely engineered to operate on nucleotides as triplets, leading to translation frameshifts. The system design would now need all kinds of error-correcting logic.

The cellular design, however, has already taken this critically important factor into account, by using binding interactions and dimensions that are just right for all the codes. Consider how epigenetic tags must be added and removed reliably, in some cases easily (histone code: see http://www.cellsignal.com/contents/resources-reference-tables/histone-modification-table/science-tables-histone) while in other cases with much more difficulty (e.g., methylation of DNA). The codon-anticodon strength plus supportive context in ribosomes ensures an interaction strong enough to ensure reliable identification but not so strong the mRNA would often stall. There are many examples of this "just right" principle that the codes implicitly rely on, such as the TF to CRE interactions, DNA-DNA base pairing strength, histone to DNA interaction, phosphorylation to proteins, and so on.

Location of synapses

Guidance of action potentials along neurons are dramatically aided by having the presynaptic terminals (where neurotransmitter molecules are released) already located at the ends of axons, strategically very near the postsynaptic neuron.

Correct range of pH, temperature, viscosity

Cells regulate their internal environment in many ways, including pH, temperature, free radical content, waste products, and viscosity. This automatically lowers the entropy with respect to a reference state lacking these controlling factors, thus circumventing the need for coded instructions to avoid the unwanted possibilities.

Cells interacting as ensembles

Striated or skeletal muscles consist of several cells that have fused together. They are excitable by neuronal stimulation, which causes them to contract, but the direction and range of possible behavior have already been established using physical constraints. For example, they do not need special instructions for how to restore the resting length since they have suitable elastic properties (Frontera and Ochala, 2014).

Discussion and Conclusions

In Part 1 (Truman, 2016) we saw that cellular processes are controlled using Boolean logic and many codes, each with their own data types, data structures, and operations. The resemblance to software used by digital computer is inescapable. Here in Part 2 we complete the picture by discussing the hardware aspects of the cellular information processing systems, necessary to instantiate the symbolic formalism and to provide physical constraints to ensure the correct outcomes.

The cell must be understood as a holistic entity. We agree with Moss's view: "I have begun to marshal evidence on behalf of the idea that cellular context as a whole is basic to the nature and continuity of living beings and is irreducible to any of its constituent parts" (Moss, 2004, p. 95). The guidance is not provided by a blueprint or detailed prescriptive language on DNA, a notion associated with the theory of preformationism prevalent in the eighteenth and nineteenth centuries (Francis, 2011, pp. 120, 126; Oyama, 2002, ch. 3). An example by Francis (2011, p. 86, 111) demonstrates this. Most (tens of millions) of epigenetic methyl tags are removed during the process of making animal eggs and sperm. Then the methylation patterns are restored in the egg and sperm cells before fertilization. Obviously, the guidance to restore the pattern must have come from an informational source other than the now methyl-free DNA. Then after fertilization, but before implantation, many of the methyl tags are removed again. In the case of plants, epigenetic tags developed during a lifetime can be stable over hundreds of generations, since epigenetic reprogramming is much less pervasive (Francis, 2011, pp. 87, 90).

Challenge to Produce One New Protein by Chance

A large variety of proteins from unrelated protein families are needed to build molecular machines (and often long RNA chains are also needed). The vast majority of random polypeptide sequences, however, are unsuitable for biological purposes, but until a new protein provides a measurable level of functionality, natural selection cannot favor that organism. One way to estimate the proportion of useful to worthless sequences is to recognize that globular proteins must first be able to fold reliably so that the various steric and electronic features are located reliably every time that protein is produced.

Douglas Axe replaced small clusters of side-chains randomly starting with a weakly functional β-lactamase domain to estimate the proportion that fold. Taking into account that about 10^4 protein fold types exist in cells, he concluded

that roughly 1 in 10^{74} domain-sized sequences fold reliably (Axe, 2004, 2010).

The literature on folding proportion was reviewed in a series of six papers in the *Journal of Creation* (Truman, 2012a) that concluded that the only unambiguous cases of folded artificial proteins produced in a lab so far had been intelligently designed. This guidance came from scientists applying a deep understanding of optimal polypeptide sizes to permit folding (considerably smaller than a median sized protein) and which amino acid patterns favor formation of necessary substructures for folding, such as α-coils and β-sheets.

The literature needs to be interpreted carefully. Polypeptide chains can form semi-structured but still basically amorphous molten globule states with various collapsed conformations, crude cooperative unfolding, and secondary structures. It is not surprising that various polypeptide sequences reveal through spectroscopy some secondary structure in some randomizing experiments, but a few random α-helices and β-sheets produced, of different sizes and locations each time a polypeptide sample clumps together, is not the same thing as a protein reliably and identically folding to a single state.

Three general approaches have been used to estimate this proportion (Truman, 2006): (1) designed "semi-random" chains of amino acids are generated to see if any show evidence of folding; (2) existing protein sequences are mutated to see how much change is tolerated; (3) sequence variability across organisms is compared. The available estimates range widely, with the examples discussed ranging between one out of 10^{44} to 10^{112} (Truman, 2006).

As an example experiment conducted at MIT, one to three codons were varied at 25 positions along a 92-residue portion of a 237 amino acid γ-repressor (Reidhaar-Olson and Sauer, 1990). For only the 92-residue portion examined, a proportion of about one out of 10^{63} sequences was estimated to be functional. In a recent study using 71 residue random-sequence proteins with overall composition tailored to resemble globular proteins, the authors commented that "a mere 5–20% of a given protein's amino acid sequence remains invariant during evolution," implying that forming proteins by chance is easy (LaBean et al., 2011).

But let us examine LaBean's estimate. Assume 5 to 20% must be invariant and every other combination in 80–95% of a protein could be acceptable. Then each kind of protein with median length 250 residues would only need to have n = 12.5 to 50 residues placed correctly, where each position would have a chance of about 1/20 being correct by chance (depending on the model assumptions, such as prebiotic amino acid proportions, or a genetic code). $(1/20)^n$ for n = 12.5 to 50 implies a proportion of only 10^{-17} to 10^{-65}. The reality, however, is that only a small proportion of proteins having the remaining 80 to 95% positions totally random would still fold reliably, so the proportion on average is surely many orders of magnitude smaller than 10^{-17}. Even small proteins such as cytochromes (only 104 residues) show identity at 26 positions for all the sequences available (Bray, 2009, p. 141), and $(1/20)^{26} = 1.5 \times 10^{-34}$.

To put the analysis in perspective, using optimistic evolutionary assumptions demonstrates that the maximum number of polypeptide candidates that could have been generated by living organisms would be about 10^{46}, spread throughout all organisms and over more than four billion years (Truman and Heisig, 2001). To produce a single new, minimally functional molecular machine, many unrelated proteins must be present at the same time, proportion, and location only to satisfy the hardware portion of the processing requirements.

Using the highest proportion of 10^{-17} from above would imply odds of obtaining a single copy of three different folded proteins as $(10^{-17})^3$ or (10^{-51}), and these must be all located together at the same time and place and somehow jointly happen to do something biologically useful. Recall that at most 10^{46} organisms ever lived. How could ribosomes, RNA and DNA polymerizes have been produced by chance?

A new complex machine such as a spliceosome would also require that the corresponding software components be in place: separating introns and exons requires the correct signals at the right place and that false positives not be spread all over the genome.

When analyzing molecular machines one should consider the size of the proteins used to create them, since larger proteins are much more difficult to form by chance. As an orientation, the median number of amino acids composing proteins has been calculated to be 361 for Eukarya, 267 for Bacteria, and 247 for Archea (Brocchieri and Karlin, 2005).

Decoders Must Be Inheritable

During mitosis, copies of the various coded data plus functioning copies of the decoding MMs all need to be replicated correctly for the daughter lineage to still function and preserve the cell type. It staggers credibility to argue this could have arisen through random mutations. For example, to replicate a pattern of 100–200 million methyl groups, specialized protein methyltransferase enzymes must already exist and be assembled at the site during creation of the new DNA (Woodward and Gills, 2012, pp. 113–114). As a second example, while a new copy of DNA is being synthesized, millions of the 8-histone spools are also manufactured for the new DNA (Woodward and Gills, 2012, p. 60).

Repair Mechanisms

Reflection shows neo-Darwinian theory (NDT) is inadequate to explain the origin of sophisticated computational machines like cells, with elaborate logic

processing, multiple codes and many very complex collaborating decoders. They are functioning von Neumann self-replicating automatons (Barbieri, 2003, p. 25). This kind of complexity is foreign to inanimate matter and until carefully set up to be adaptable and self-correcting would quickly collapse as a viable system. More than 10,000 DNA damage events occur daily in every cell in the human body, and if left to itself, a malfunctioning (cancerous) cell would spread the damage. We survive because our cells possess about 10 different DNA repair pathways, designed to handle particular lesions (Mullins et al., 2015).

Error correcting processes—such as double-strand break (DSB) (Brissett and Doherty, 2009) and nucleotide excision repair (NER) (Harfe and Jinks-Robertson, 2000)—probe the genome for endogenous and exogenous damage, and must be able to recognize special patterns that signal damage, which incidentally represent yet another code. Codes and processing equipment are ubiquitous in cellular repair. Noncoding RNAs are expressed in response to damage (Chowdhury et al., 2013) and a number of phosphorylation, acetylation, methylation, ubiquitylation and chromatin-remodeling codes are used to demarcate the lesion site. These provide the informational input to first assemble and then break down the required repair machines (Gospodinov and Herceg, 2013; Hitomi et al., 2007; Marteijn et al., 2014; Panier and Durocher, 2013). Damaged DNA is often mobilized to subnuclear repair centers (Shapiro, 2011 p. 46; Bekker-Jensen and Mailand, 2010; Lisby and Rothstein, 2004), another example of the dynamic interaction between informational parameters and dynamic hardware processors.

Genome Truncation

A final reason we'll mention for the irrelevance of NDT involves the requirement to start simple and increase complexity. Many complex molecular machines, such as spliceosomes, are only found in eukaryotes, which supposedly arose long after prokaryotes, so the theory must explain the origin of dramatic innovations and much larger genomes. Eukaryotes have far longer generation times and much lower population sizes, so considerably fewer than 10^{46} complex organisms would have been available to produce a multitude of novel proteins *de novo*. Needed are not trivially different protein sequences but proteins distributed among very different protein families. At least 4909 protein families are used in biology (Punta et al., 2012).

NDT is all about competitive replication from common ancestors. This poses a big problem. For small prokaryote organisms having around 1000 genes in the distant past, individuals expressing unnecessary genes would be disadvantaged, and streamlined competing individuals or species would soon out-populate them. The components of new molecular machines (without which the new codes would be worthless) would have to offer a net and immediate benefit above the disadvantages of longer chromosome replication times and greater energy plus material demands.

The new systems could not develop step-wise, waiting millions of generations for future fortuitous mutations to eventually generate a minimal level of selectively measurable functionality. The only way to compete against sister species in that immediate environment would be to truncate genomes, by eliminating whatever genetic material is not needed at the moment. Bacteria constantly eliminate DNA, and genome streamlining is a well-known phenomenon (Wagner, 2005, 2007; Truman and Terborg, 2008a, 2008b).

Neo-Darwinism Fails to Explain the Origin of Logic Processing

Neo-Darwinian theory remains the dominant naturalist evolutionary theory and assumes all extant life evolved from a common ancestor, despite ever-increasing gaps in credibility. Historically it developed from a philosophical desire to deny the activity of a creator. Shapiro, from the University of Chicago, pointed out that "this insistence on randomness and accident is not surprising. It springs from a determination in the 19th and 20th Centuries by biologists to reject the role of a supernatural agent in religious accounts of how diverse living organisms originated" (Shapiro, 2011, p. 1).

Müller, from the University of Vienna, summarizes well where the theory now stands:

> In the neo-Darwinian world, the motive factor for morphological change is natural selection, which can account for the modification and loss of parts. But selection has no innovative capacity: it eliminates or maintains what exists.... The inability of evolutionary theory to account for phenotypic organization has been recognized by numerous authors, with regards to both biochemical and morphological evolution. (Müller, 2003, p. 51)

In general, whenever we come across the terms "convergent evolution," "genetic piracy," or "exaptation," we will discover a failure of neo-Darwinian theory and in most cases novel logic-processing elements which produce the same kind of biological feature. Müller goes into a detailed and very illuminating analysis of how "homology of a character in different species does not mean this character is generated by identical developmental processes, is controlled by the same genes" (Müller, 2003, p. 56) and admits to "the growing tendency to reject homology as a serious scientific topic" (p. 56). In the same book, Britten concludes that the prospects of finding a tree of life has become ever more remote, as more genomic sequences of microorganism have become available (Britten, 2003, p. 79). Of what value is a theory that fails to provide research

guidance or consistent interpretative methods?

In discussing the case of the octopus and the human camera eye, Kirschner and Gerhart (2005, p. 240) point out that the eyes derive from different tissues by different developmental means and are based on completely different phototransduction circuitries. Müller makes the same point (Müller, 2003, p. 57). Nevertheless, both use eye designs having the same pigment (rhodopsin) for photoreception, and both send electrical signals to the brain. Most now agree that a common ancestry for this kind of eye makes no sense, especially since organisms similar to the octopus do not share this feature, so the word "convergence" is used, which actually explains nothing whatsoever.

Remarkably, Kirschner and Gerhart (2005, pp. 253–254) claim tubulin and actin proteins must have been present in a common ancestor of prokaryotes and eukaryotic cells on the basis of having virtually the same three-dimensional structures even though they have almost zero sequence resemblance. The evolutionary interpretive methodology apparently permits picking and discarding what is due to common ancestry versus convergence on the basis of whichever seems to cause the least difficulties.

Having recognized that assuming common ancestry has provided him with no guidance, Müller continues asking the right questions and comes up with a sensible conclusion. "I propose that homology is not merely a concept or a conceptual tool, as it is often understood, but rather the manifestation of morphological organization processes. It thus represents a major, unresolved problem in evolutionary biology" (Müller 2003, p. 51).

The words "evolution" and "natural selection" have been given unbounded creative powers. Müller is now assigning a similar role to the word "homology." It is a placeholder for an organizing principle that has been instantiated in many ways but with a purposeful outcome which humans are able to recognize. "Homology is the manifestation of an ordering principle in morphological evolution … This legitimate and strangely neglected scientific question lies at the center of the phenomenon of organismal evolution: it is not satisfactorily answered by current evolutionary theories" (Müller, 2003, p. 58).

What could these organizational processes be but a judicious application and reapplication of the kinds of common software and hardware informational principles we have been discussing?

NDT offers no solution to the existence of logic processing found throughout cells or the putative dramatic increase in complexity over time. In fact, it must assume a miracle again and again to permit conceptual starting points. Systematic genomic comparison studies are leading to the view (Lerat et al., 2004; Lerat et al., 2005) that a core of about 100 different genes are indispensable in all organisms. Anything less than this would not work. These core genes plus a few hundred others that may be somewhat exchangeable would have to be regulated and operational in a membrane protected up-and-running system. A few hundred genes' worth of DNA on their own in a fluid environment will perform no biochemistry; they will simply be a dead polymer like a wet piece of plastic. This poses an insurmountable barrier for natural processes to produce a minimal organism able to survive and reproduce reliably.

Excluding a determined philosophically prior commitment, there is no solid scientific basis to believe unguided natural processes could or have created what hundreds of thousands of scientists cannot. As Bray pointed out, "Not only are biologists incapable at present of manufacturing the enzymes, membranes, and organelles needed for these processes; they still do not fully understand how they work or are put together" (Bray, 2009, p. 207).

We agree with Abel that "formalism not only describes, but preceded, prescribed, organized and continues to govern and predict Physicality" (Abel, 2011, p. 325). Cells, like all information-processing equipment, require a cognizant designer.

Acknowledgments

We wish to thank Dr. Peer Terborg and Dr. Niko Winkler for many fruitful discussions, Dr. Michael Heisig for help obtaining much of the literature cited here, Jonathan Bartlett for critical analysis and excellent suggestions, Wort und Wissen in Germany for many conferences and opportunities to discuss these topics, and Creation Ministries International for years of fruitful collaboration in this area.

References

Abel, D.L. (editor). 2011. *The First Gene: The Birth of Programming, Messaging and Formal Control*. LongView Press, New York, NY.

Adams, R.L, and R.J. Bird. 2009. Review article: Coagulation cascade and therapeutics update: relevance to nephrology. Part 1: Overview of coagulation, thrombophilias and history of anticoagulants. *Nephrology* 14:462–470.

Ahlquist, P. 2002. RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science* 296(5571):1270–1273.

Alberts, B. 1998. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92:291.

Allan, V.J. 2011. Cytoplasmic dynein. *Biochemical Society Transactions* 39:1169–1178.

Amm, I., T. Sommer, and D.H. Wolf. 2014. Protein quality control and elimination of protein waste: the role of the ubiquitin-proteasome system. *Biochimica et Biophysica Acta* 1843(1):182–196.

Anderson, P., and N. Kedersha. 2006. RNA granules. *The Journal of Cell Biology* 172(6):803–808.

Anderson, P., and N. Kedersha. 2009. RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nature Reviews Molecular Cell Biology* 10:430–436.

Ashe, H.L., and J. Briscoe. 2006. The interpretation of morphogen gradients. *Development* 133:385–394.

Axe, D.D. 2004. Estimating the prevalence of protein sequences adopting functional enzyme folds. *Journal of Molecular Biology* 341:1295–1315.

Axe, D.D. 2010. The Case Against a Darwinian Origin of Protein Folds. *BioComplexity* 1:1–12.

Baker, T.A., and R.T. Sauer. 2012. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochimica et Biophysica Acta* 1823:15–28.

Ball, L.G., M.D. Hanna, A.D. Lambrecht, B.A. Mitchell, B. Ziola, et al. 2014. The Mre11-Rad50-Xrs2 complex is required for yeast DNA postreplication repair. *PLoS ONE* 9(10):e109292.

Balzani, V., A. Credi, F.M. Raymo, and J.F. Stoddart. 2000. Artificial molecular machines. *Angewante Chemie International Edition* 39:3348–3391.

Bandyra, K.J., M. Bouvier, A.J. Carpousis, and B.F. Luisi. 2013. Review. The social fabric of the RNA degradosome. *Biochimica et Biophysica Acta* 1829:514–522.

Barbieri, M. 2003. *The Organic Codes. An Introduction to Semantic Biology*. Cambridge University Press, Cambridge, UK.

Bar-Nun, S., M.H. Glickman. 2012. Proteasomal AAA-ATPases: structure and function. *Biochimica et Biophysica Acta (BBA)—Molecular Cell Research* 1823(1):67–82.

Baudat, F., Y. Imai, and B. de Massy. 2013. Meiotic recombination in mammals: localization and regulation. *Nature Reviews Genetics* 14:794–806.

Bekker-Jensen, S., and N. Mailand. 2010. Assembly and function of DNA double-strand break repair foci in mammalian cells. *DNA Repair (Amst.)* 9:1219–1228.

Berget, S.M. 1995. Exon recognition in vertebrate splicing. *Journal of Biological Chemistry* 270:2411–2414.

Besche, H.C., P. Andreas, and A. L. Goldberg. 2009. Getting to first base in proteasome assembly. *Cell* 138:25–28.

Bhattacharyya, S., Y. Houqing, M. Carsten, and A. Matouschek. 2014. Regulated protein turnover: snapshots of the proteasome in action. *Nature Reviews Molecular Cell Biology* 15:122–133.

Boucher, L.E., and J. Bosch. 2015. The apicomplexan glideosome and adhesins—structures and function. *Journal of Structural Biology* 190:93–114.

Bray, D. 2009. *Wetware: A Computer in Every Living Cell*. Yale University Press, New Haven, CT.

Brick, K., F. Smagulova, P. Khil, R.D. Camerini-Otero, and G.V. Petukhova. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485:642–645.

Brini, M., T. Cali, D. Ottolini, and E. Carafoli. 2013. The plasma membrane calcium pump in health and disease. *FEBS Journal* 280 (supplement 21):5385–5397.

Brissett, N.C., and A.J. Doherty. 2009. Repairing DNA double-strand breaks by the prokaryotic non-homologous end-joining pathway. *Biochemical Society Transactions* 37:539–545.

Britten, R.J. 2003. Only details determine. In Müller, G.B., and S.A. Newmann (editors), *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*, pp. 75–86. MIT Press, Cambridge, MA.

Brocchieri, L., and S. Karlin. 2005. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Research* 22(10):3390–3400.

Brown, W.E., K.H. Stump, and W.S. Kelley. 1982. Escherichia coli DNA polymerase I. *Journal of Biological Chemistry* 257(4):1965–1972.

Chatterjee, S., S. Chaudhury, A.C. McShan, K. Kaur, and R.N. De Guzman. 2013. Structure and biophysics of type III secretion in bacteria. *Biochemistry* 52(15):2508–2517.

Cheeseman, I.M., and A. Desai. 2008. Molecular architecture of the kinetochore-microtubule interface. *Nature Reviews Molecular Cell Biology* 9:33–46.

Chen, S.S., and J.R. Williamson. 2013. Characterization of the ribosome biogenesis landscape in E. coli using quantitative mass spectrometry. *Molecular Biology* 425(4):767–779.

Chen, W., and M.J. Moore. 2014. Spliceosome disorder and dynamics defined. *Current Opinion in Structural Biology* 24:141–149.

Chowdhury, D., Y.E. Choi, and M.E. Brault. 2013. Charity begins at home: non-coding RNA functions in DNA repair. *Nature Reviews Molecular Cell Biology* 14:181–189.

Costa, T.R.D., C. Felisberto-Rodrigues, A. Meir, M.S. Prevost, A. Redzej, M. Trokter, and G. Waksman. 2015. Secretion systems in gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology* 13:343–359.

Cotterill, S., and S. Kearsey. 2014. Eukaryotic DNA polymerases. Encyclopedia of Life Sciences, http://www.els.net/WileyCDA/ElsArticle/refId-a0001045.html

Cournac, A., and J. Plumbridge. 2013. DNA looping in prokaryotes: experimental and theoretical approaches. *Journal of Bacteriology* 195(6):1109–1119.

Doherty, G.J., and H.T. McMahon. 2008. Mediation, modulation, and consequences of membrane-cytoskeleton interactions. *Annual Review of Biophysics* 37:65–95.

Douzi, B., A. Filloux, and R. Voulhoux. 2012. On the path to uncover the bacterial type II secretion system. *Philosophical Transactions of the Royal Society B* 367:1059–1072.

Ducrest, A.L., H. Szutorisz, J. Lingner, and M. Nabholz. 2002. Regulation of the human telomerase reverse transcriptase gene. *Oncogene* 21:541–552.

Dustin, M.L. 2014. What counts in the immunological synapse? *Molecular Cell Review* 54:255–262.

Ebright R.H. 2000. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *Journal of Molecular Biology* 304(5):687–698.

Erbas-Cakmak, S., D.A. Leigh, C.T. McTernan, and A.L. Nussbaumer. 2015. Artificial molecular machines. *Chemical Reviews* 115(18):10081–10206.

Farmer, S., P.A. San-Segundo, and L. Aragon. 2011. The Smc5–Smc6 complex is required to remove chromosome junctions in meiosis. *PLoS ONE* 6(6):e20948.

Federley, R.G., and L.J. Romano. 2010. DNA polymerase: structural homology, conformational dynamics, and the effects of carcinogenic DNA adducts. *Journal of Nucleic Acids* 2010: Article ID 457176.

Fletcher, D.A., and R.D. Mullins. 2010. Review article cell mechanics and the cytoskeleton. *Nature* 463:485–492.

Fink, A.I. 1999. Chaperone-mediated protein folding. *Physiological Reviews* 79(2):425–442.

Francis, R.C. 2011. *Epigenetics: How Environment Shapes our Genes*. W.W. Norton & Company, New York, NY.

Frénal, K., V. Polonais, J-B. Marq, R. Stratmann, J. Limenitakis, and D. Soldati-Favre. 2010. Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host & Microbe* 8:343–357.

Frenkel-Morgenstern, M., et al. 2012. Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. *Genome Research* 22(7):1231–1242.

Frontera, W.R., and J. Ochala. 2014. Skeletal Muscle: a Brief Review of Structure and Function, *Calcified Tissue International* 96(3):183–195.

Fuchs, Y., and H. Steller. 2011. Programmed cell death in animal development and disease. *Cell* 147:742–758.

Garcia-Diaz, M., and K. Bebenek. 2007. Multiple functions of DNA polymerases. *Critical Reviews in Plant Science* 26(2):105–122.

Gazzaniga, M.S., R.B. Ivry, and G.R. Mangun. 2009. *Cognitive Neuroscience: The Biology of the Mind*, 3rd edition. W. W. Norton & Company, Inc., New York, NY.

Głogocka, D., M. Przybyło, and M. Langner. 2015. Molecular machines—a new dimension of biological sciences. *Cellular and Molecular Biology Letters* 20(2):248–264.

Gospodinov, A., and Z Herceg. 2013. Shaping chromatin for repair. *Mutation Research* 752:45–60.

Hanson, P.I., and S.W. Whiteheart. 2005. AAA+ proteins: have engine, will work. *Nature Reviews Molecular Cell Biology* 6:519–529.

Harfe, B., and S. Jinks-Robertson. 2000. DNA mismatch repair and genetic instability. *Annual Review of Genetics* 34:359–399.

Haupts, U., J. Tittor, and D. Oesterhelt. 1999. Closing in on bacteriorhodopsin: progress in understanding the molecule. *Annual Review of Biophysics and Biomolecular Structure* 28:367–399.

Herai, R.H., and M.E. Yamagishi. 2010. Detection of human interchromosomal trans-splicing in sequence databanks. *Briefings in Bioinformatics* 11(2):198–209.

Herrmann, H., H. Bär, L. Kreplak, S.V. Strelkovl, and U. Aebi. 2007. Intermediate filaments: from cell architecture to nanomechanics. *Nature Reviews Molecular Cell Biology* 8:562–573.

Hirano, T. 2012. Condensins: universal organizers of chromosomes with diverse functions. *Genes & Development* 26:1659–1678.

Hitomi, K., S. Iwai, and J.A. Tainer. 2007. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst.)* 6:410–428.

Hou, Y.M., and P. Schimmel. 1988. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* 333:140–145.

Huber, F., J. Schnauß, S. Rönicke, P. Rauch, K. Müller, C. Fütterer, and J. Käsa. 2013. Emergent complexity of the cytoskeleton: from single filaments to tissue. *Advances in Physics* 62(1):1–112.

Hübscher U., H.-P. Nasheuerb, and J.E. Syväoja. 2000. Eukaryotic DNA polymerases, a growing family. *Trends in Biochemical Sciences* 25(3):143–147.

Ishihama, A. 2000. Functional modulation of Escherichia coli RNA polymerase. *Annual Review of Microbiology* 54:499–518.

Iwasaki, A., and R. Medzhitov. 2015. Control of adaptive immunity by the innate immune system. *Nature Immunology* 16:343–353.

Jankowsky, E. 2011. RNA helicases at work: binding and rearranging. *Trends in Biochemical Sciences* 36(1):19–29.

Jonkers, I., and J.T. Lis. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* 16:167–177.

Kiebler, M.A., and G.J. Bassell. 2006. Neuronal RNA granules: movers and makers. *Neuron* 51:685–690.

King, M.W. 2015. Introduction to pyruvate metabolism and the TCA cycle. http://themedicalbiochemistrypage.org/tca-cycle.php

Kirschner, M.W., and J.C. Gerhart. 2005. *The Plausibility of Life: Resolving Darwin's Dilemma*. Yale University Press, New Haven, CT.

Koehler, C.M., S. Merchant, and G. Schatz. 1999. How membrane proteins travel across the mitochondrial intermembrane space. *Trends in Biochemical Sciences* 24(11):428–432.

Kunkel, T.A., and K. Bebenek. 2000. DNA replication fidelity. *Annual Review of Biochemistry* 69:497–529.

LaBean, T.H., T.R. Butt, S.A. Kauffman, and E.A. Schultes. 2011. Protein folding absent selection. *Genes* 2:608–626.

Lasda, E.L., and T. Blumenthal. 2011. Trans-splicing. *Wiley Interdisciplinary Reviews. RNA* 2(3):417–434.

Lehn, J-N. 2002. Supramolecular chemistry and self-assembly. Toward self-organization and complex matter. *Science* 295:2400–2403.

Lerat, E., V. Daubin, and N.A. Moran. 2004. *From gene trees to organismal phylogeny in* prokaryotes: the case of the g-Proteobacteria. *PLoS Biology* 1(1):101–109.

Lerat, E., V. Daubin, H. Ochman, and N.A. Moran. 2005. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biology* 3(5):0807–0814.

Lisby, M., and R. Rothstein. 2004. DNA damage checkpoint and repair centers. *Current Opinion in Cell Biology* 16:328–334.

Lisby, M., and R. Rothstein. 2005. Localization of checkpoint and repair proteins in eukaryotes. *Biochimie* 87:579–589.

Lodish, H. et al. 2000. *Molecular Cell Biology*. W. H. Freeman and Company, New York, NY.

Long, Q., K. Yang, and Q. Yang. 2015. Regulation of mitochondrial ATP synthase in cardiac pathophysiology. *American Journal of Cardiovascular Disease* 5(1):19–32.

Lue, N.F. 2009. Closing the feedback loop: how cells "count" telomere-bound proteins. *Molecular Cell Previews* 33:413–414.

Luskin, C. 2010. Molecular machines in the cell. Discovery Institute. http://www.discovery.org/a/14791.

Lycklama a Nijeholt, J.A., and A.J.M. Driessen. 2012. The bacterial Sec-translocase: structure and mechanism. *Philosophical Transactions of the Royal Society B* 367:1016–1028.

Maddox, P.S., and A.M. Ladouceur. 2015. Concentrating on the mitotic spindle. *Journal of Cell Biology* 210(5):691–693.

Maida, Y., and K. Masutomi. 2011. RNA-dependent RNA polymerases in RNA silencing. *Biological Chemistry* 392(4):299–304.

Marteijn, J.A., H. Lans, W. Vermeule, and J.H.J. Hoeijmakers. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology* 15:465–481.

Marx, A., A. Hoenger, and E. Mandelkow. 2009. Structures of kinesin motor proteins. *Cell Motility and the Cytoskeleton* 66(11):958–966.

Matera, A.G., and Z. Wang. 2014. A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology* 15:108–121.

Mavroidis C., A. Dubey, and M.L. Yarmush. 2004. Molecular machines. *Annual Review of Biomedical Engineering* 6:363–395.

Mayer, M.P. 2013. Hsp70 chaperone dynamics and molecular mechanism. *Trends in Biochemical Sciences* 38(10):507–514.

Mayer, M.P., and B. Bukau. 2005. Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences* 62:670–684.

McCulloch, S.D., and T.A. Kunkel. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Research* 18(1):148–61.

Mehta, G.D., S.M.A. Rizvi, and S.K. Ghosh. 2012. Cohesin: a guardian of genome integrity. *Biochimica et Biophysica Acta (BBA)—Molecular Cell Research* 1823(8):1324–1342.

Melnik, S., B. Deng, A. Papantonis, S. Baboo, I.M. Carr, and P.R. Cook. 2011. The proteomes of transcription factories containing RNA polymerases I, II or III. *Nature Methods* 8:963–968.

Merendino, A.M., F. Bucchieri, C. Campanella, V. Marcianò, A. Ribbene, S. David, et al. 2010. Hsp60 is actively secreted by human tumor cells. *PLoS ONE* 5(2):e9247.

Meuleman, W., et al. 2013. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Research* 23(2):270–280.

Michaeli, S. 2011. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiology* 6(4):459–474.

Michelet, L., M. Zaffagnini, S. Morisse, F. Sparla, M.E. Pérez-Pérez, F. Francia, A. Danon, C.H. Marchand, S. Fermani, P. Trost, and S.D. Lemaire. 2013. Redox regulation of the Calvin-Benson cycle: something old, something new. *Frontiers in Plant Science* 4(470):1–21.

Mignea, A. 2014. Developing insights into the design of the simplest self-replicator and its complexity, Parts 1–3. In Bartlett, J., D. Halsmer, and M.R. Hall, M.R. (editors), *Engineering and the Ultimate*, pp. 169–219. Blyth Institute Press, Broken Arrow, OK.

Misteli, T., and E. Soutoglou. 2009. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature Reviews Molecular Cell Biology* 10(4):243–254.

Mitchell, J.A., and P. Fraser. 2008. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Development* 22(1):20–25.

Montilla-Martinez, M., et al. 2015. Distinct pores for peroxisomal import of PTS1 and PTS2 proteins. *Cell Reports* 13:1–9.

Moss, L. 2004. *What Genes Can't Do*. MIT Press, Cambridge, MA.

Müller, G.B. 2003. Homology: the evolution of morphological organization. In Müller, G.B., and S.A. Newmann (editors), *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*, pp. 51–70. MIT Press, Cambridge, MA.

Müller, G.B., and S.A. Newmann (editors). 2003. *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*. MIT Press, Cambridge, MA.

Mullins, E.A., R. Shi, Z.D. Parsons, P.K. Yuen, S.S. David, Y. Igarashi, and B.F. Eichman. 2015. The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature* 527:254–258.

Nakamoto, R.K., J.A.B Scanlon, and M.K. Al-Shawi. 2008. The rotary mechanism of the ATP synthase. *Archives of Biochemistry and Biophysics* 476:43–50.

Nilsen, T.W. 2003. The spliceosome: the most complex macromolecular machine in the cell? *BioEssays* 25:1147–1149.

Nitiss, J.L. 2009. DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews Cancer* 9(5):327–337.

Norcum, M.T., C.L. Wolfe, and J.A. Warrington. 2005. Three-dimensional working model of the multienzyme aminoacyl-tRNA synthetase complex determined by computational microscopy. *Microscopy and Microanalysis* 11 (Supplement 2):164–165.

Nureki, O., D.G. Vassylyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M. Konno, T.L. Hendrickson, P. Schim-

mel, and S. Yokoyama. 1998. Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* 280(5363):578–582.

Osborne, C., L. Chakalova, K.E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J.A. Mitchell, S. Lopen, W. Reik, and P. Fraser. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genetics* 36(10):1065–1071.

O'Sullivan, R.J., and J. Karlseder. 2010. Telomeres: protecting chromosomes against genome instability. *Nature Reviews Molecular Cell Biology* 11:171–181.

Outwin, E.A., A. Irmisch, J.M. Murray, and M.J. O'Connell. 2009. Smc5-Smc6-dependent removal of cohesin from mitotic chromosomes. *Molecular and Cell Biology* 29(16):4363–4375.

Oyama, S. 2002. *The Ontogeny of Information: Developmental Systems and Evolution*. 2nd edition. Duke University Press, Durham, NC.

Panier, S., and D. Durocher. 2013. Push back to respond better: regulatory inhibition of the DNA double-strand break response. *Nature Reviews Molecular Cell Biology* 14:661–672.

Panwar, B., and G.P.S. Raghava. 2010. Prediction and classification of aminoacyl tRNA synthetases using PROSITE domains. *BMC Genomics* 11:507.

Pazour, G.J., N. Agrin, J. Leszyk, and G.B. Witman. 2005. Proteomic analysis of a eukaryotic cilium. *Journal of Cell Biology* 170(1):103–113.

Pike, L.J. 2009. The challenge of lipid rafts. *Journal of Lipid Research*. 50(Supplement):S323-S328.

Portt, L., G. Norman, C. Clapp, M. Greenwood, and M.T. Greenwood. 2011. Anti-apoptosis and cell survival: a review. *Biochimica et Biophysica Acta* 1813:238–259.

Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A Bateman, and R.D. Finn. 2012. The Pfam protein families database. *Nucleic Acids Research* 40(Database issue):D290-D301.

Rastogi, R.P, and R.P. Sinha. 2009. Apotosis: molecular mechanisms and pathogenicity. *EXCLI Journal* 8:155–181.

Razin, S.V., A.A. Gavrilov, A. Pichugin, M. Lipinski, O.V. Iarovaia, and Y.S. Vassetzky. 2011. Transcription factories in the context of the nuclear and genome organization. *Nucleic Acids Research* 39(21):9085–9092.

Reidhaar-Olson, J.F., and R.T. Sauer. 1990. Functionally acceptable substitutions in two α-helical regions of λ repressor. *Proteins: Structure, Function and Genetics* 7:306–316.

Rieder, D., Z. Trajanoski, and J.G. McNally. 2012. Transcription factories. *Frontiers in Genetics* 3(221):1–12.

Sadaie, M., T. Naito, and F. Ishikawa. 2003. Stable inheritance of telomere chromatin structure and function in the absence of telomeric repeats. *Genes and Development* 17(18):2271–2282.

Sarfati, J. 1999. Decoding and editing designs: double-sieve enzymes. *Journal of Creation* 13(1):5–7.

Schechter, A.N. 2008. Hemoglobin research and the origins of molecular medicine. *Blood* 112(10):3927–3938.

Schenten, D., and R. Medzhitov. 2011. The control of adaptive immune responses by the innate immune system. *Advances in Immunology* 109:87–124.

Schmidt, O., N. Pfanner, and C. Meisinger. 2010. Mitochondrial protein import: from proteomics to functional mechanisms. *Nature Reviews Molecular Cell Biology* 11:655–667.

Shapiro, J.A. 2006. Genome informatics: the role of DNA in cellular computations. *Biological Theory* 1(3):288–301.

Shapiro, J.A. 2011. *Evolution: A View from the 21st Century*. FT Press Science, Upper Saddle River, NJ.

Shapiro, J.A. 2013. How life changes itself: the read-write (RW) genome. *Physics of Life Reviews* 10:287–323.

Shapiro, J.A., and R. von Sternberg. 2005. Why repetitive DNA is essential to genome function. *Biological Reviews* 80:1–24.

Simons, K., and R. Ehehalt. 2002. Cholesterol, lipid rafts, and disease. *Journal of Clinical Investigation* 110(5):597–603.

Sommers, J.A., A.N. Suhasini, and R.M. Brosh Jr. 2015. Protein degradation pathways regulate the functions of helicases in the DNA damage response and maintenance of genomic stability. *Biomolecules* 5:590–616.

Spector, D.L. 2001. Nuclear domains. *Journal of Cell Science* 114:2891–2893.

Staley, J.P., and J.L. Woolford Jr. 2009. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Current Opinion in Cell Biology* 21(1):109–118.

Steinberg, S.F. 2008. Structural basis of protein kinase C isoform function. *Physiological Reviews* 88(4):1341–1378.

Sugiyama, T., et al. 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proceedings of the National Academy of Sciences USA* 102(1):152–157.

Sullivan, B.A., M.D. Blower, and G.H. Karpen. 2001. Determining centromere identity: cyclical stories and forking paths. *Nature Reviews Genetics* 2:584–596.

Sweeny, H.L., and A. Houdusse. 2010. Structural and functional insights into the myosin motor mechanism. *Annual Review of Biophysics* 39:539–557.

Trifonov, E.N. 2011. Thirty years of multiple sequence codes. *Genomics Proteomics Bioinformatics* 9(1–2):1–6.

Truman, R. 2006. Searching for needles in a haystack. *Journal of Creation* 20(2):90–99.

Truman, R. 2012a. The proportion of polypeptide chains which generate native folds—part 6: extracting from random sequences. *Journal of Creation* 26(1):86–93.

Truman, R. 2012b. Information theory—part 1—overview of key ideas. *Journal of Creation* 26(3):101–106.

Truman, R. 2012c. Information theory—

part 2: weaknesses in current conceptual frameworks. *Journal of Creation* 26(3):107–114.

Truman, R. 2012d. Information theory—part 3: introduction to coded information systems. *Journal of Creation* 26(3):115–119.

Truman, R. 2013. Information theory—part 4: fundamental theorems of coded information systems theory. *Journal of Creation* 27(1):71–77.

Truman, R. 2015. Nylon-eating bacteria—part 4: interpretation according to coded information system theory. *Journal of Creation* 29(3):46–52.

Truman, R. 2016. Cells as information processors Part 1. Formal Software Principles. *CRSQ* 52:275–308.

Truman, R., and M. Heisig. 2001. Protein families: chance or design? *Journal of Creation* 15(3):115–127.

Truman, R., and P. Terborg. 2008a. Genome truncation vs mutational opportunity: can new genes arise via gene duplication? Part 1. *Journal of Creation* 22(1):99–110.

Truman, R., and P. Terborg. 2008b. Genome truncation vs mutational opportunity: can new genes arise via gene duplication? Part 2. *Journal of Creation* 22(1):111–119.

Turing, A.M. 1950. Computing machinery and intelligence. *Mind* 59:433–460.

Verdel, A., et al. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303(5658):672–676.

Verdel, A., et al. 2009. Common themes in siRNA-mediated epigenetic silencing pathways. *International Journal of Developmental Biology* 53(2–3):245–257.

Wagner, A. 2005. Energy constraints on the evolution of gene expression. *Molecular Biology and Evolution* 22(6):1365–1374.

Wagner, A. 2007. Energy costs constrain the evolution of gene expression. *Journal of Exploratory Zoology Part B: Molecular and Development Evolution* 308B: 322–324.

Walhout, M., M. Vidal, and J. Dekker (editors). 2013. *Handbook of Systems Biology: Concepts and Insights*. Academic Press, London, UK.

Wang, J., and M.A. Maldonado. 2006. The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. *Cellular & Molecular Immunology* (4):255–261.

Ward, A.J., and T.A. Cooper. 2010. The pathobiology of splicing. *Journal of Pathology* 220(2):152–163.

Wikström, M. 2004. Cytochrome c oxidase: 25 years of the elusive proton pump. *Biochimica et Biophysica Acta (BBA)—Bioenergetics* 1655:241–247.

Wilson, D.N., and J.H. Doudna Cate. 2012. The structure and function of the eukaryotic ribosome. *Cold Spring Harb Perspectives in Biology* 4:a011536.

Woodward, T.E., and J.P. Gills. 2012. *The Mysterious Epigenome: What Lies Beyond DNA*. Kregel Publications, Grand Rapids, MI.

Wu, S., W.A. Beard, L.G. Pedersen, and S.H. Wilson. 2014. Structural comparison of DNA polymerase architecture suggests a nucleotide gateway to the polymerase active site. *Chemical Reviews* 114(5):2759–2774.

Xie, L., T. Altindal, S. Chattopadhyay, and X-L. Wu. 2011. Bacterial flagellum as a propeller and as a rudder for efficient chemotaxis. *Proceedings of the National Academy of Sciences* 108(6):2246–2251.

Yadavalli, S.S., and M. Ibba. 2012. Quality control in aminoacyl-tRNA synthesis: its role in translational fidelity. *Advances in Protein Chemistry and Structural Biology* 86:1–43.

Yamada, S., K. Ohta, and T. Yamada. 2013. Acetylated Histone H3K9 is associated with meiotic recombination hotspots, and plays a role in recombination redundantly with other factors including the H3K4 methylase Set1 in fission yeast. *Nucleic Acids Research* 41:3504–3517.

Yoshikawa, S., K. Muramoto, and K. Shinzawa-Itoh. 2011. Proton-pumping mechanism of cytochrome c oxidase. *Annual Review of Biophysics* 40:205–223.

Yoshikawa, S., and A. Shimada. 2015. Reaction mechanism of cytochrome c oxidase. *Chemical Reviews* 115(4):1936–1989.