

Genes without prominence: a reappraisal of the foundations of biology

Arto Annala

Department of Biosciences, Department of Physics

POB 64, Gustaf Hällströmin katu 2,

FI-00014 University of Helsinki, Finland

Keith Baverstock

Department of Environmental Science

POB 1627, Kuopio Campus

FI-70211 University of Eastern Finland

Kuopio, Finland

Abstract

The sequencing of the human genome raises two intriguing questions: why has the prediction of the inheritance of common diseases from the presence of abnormal alleles proved so unrewarding in most cases and, how can some 25,000 genes generate such a rich complexity evident in the human phenotype? It is proposed that light can be shed on these questions by viewing evolution and organisms as natural processes contingent on the 2nd law of thermodynamics, equivalent to the principle of least action in its original form. Consequently, natural selection acts upon variation in any mechanism that consumes energy from the environment, rather than on genetic variation.

According to this tenet cellular phenotype, represented by a minimum free energy attractor state comprising active gene products, has a causal role in giving rise, by a self-similar process of cell to cell interaction, to morphology and functionality in organisms, which in turn, by a self-similar process entailing Darwin's "proportional numbers", are influencing their ecosystems.

Thus, genes are merely a means of specifying polypeptides: those that serve free energy consumption in a given surroundings contribute to cellular phenotype as determined by the phenotype. In such natural processes everything depends on everything else and phenotypes are emergent properties of their systems.

Key words: 2nd law of thermodynamics; principle of least action; free energy; entropy; bound energy; attractors; epigenetics; evolution.

Introduction

The sequencing of the human genome, completed in 2001, implied that abnormal alleles could be associated with common disease traits. In the event, 12 years on, this promise has not been fulfilled. The misnamed “missing heritability” has triggered the search for the “hidden” allele quality that would be responsible for these traits; while genome wide association studies (GWAS) have uncovered “*hundreds of common variants whose allele frequencies are statistically correlated with various illnesses and traits ...the vast majority ... have no established biological relevance to disease or clinical utility for prognosis or treatment*” and so now whole genome sequencing is held out as the answer [1]. However, studies of identical twin pairs, which allow outcomes from two identical genome sequences to be compared, show that, for the majority of common diseases, knowing the causes of death or disease history of one twin gives only marginal guidance as the causes of death or disease suffered by the other [2]. Against this uncertain background plans are being put in place to sequence the whole genomes of large numbers of individuals (tens of millions) in pursuit of personalised medicine¹.

More fundamentally, biology has no substantive answer to the question “from where does the phenotypic complexity of higher mammals derive?”. What might be called the “genomic input” for human cells, in terms of numbers of gene coding sequences, splicing potential, diversity in peptide folding and measured interactome size, can be stretched to a factor of a few thousand compared to bacterial cells, but the output in terms of phenotypic function is vastly greater. This raises the question: “are these components of the input really the most relevant ones?”. Take for example the round worm, *Caenorhabditis elegans*, with about 20,000 gene coding sequences and, on average, about 5 exons per sequence, compared to the human cell, with possibly 25,000 gene coding sequences and about 8 exons/coding sequence, i.e., almost the same,

but the human cell giving rise to organisms of vastly greater phenotypic complexity than roundworms. Of course there are differences: the human genome is 30 times longer than that of *C. elegans*, but are there sound theoretical reasons for supposing that this makes the grand difference?

It is proposed here that by basing biology on the physics of dissipative systems governed by the universal law of nature, i.e., the second law of thermodynamics (hereinafter the 2nd law), insight into the origin of phenotypic complexity can be derived, which in turn sheds light on the “missing heritability” issue. Moreover, the viewpoint allows us to evaluate the prospects for personalised medicine based on whole genome sequencing. We emphasize that no new science is involved in what follows; simply the consequences of the basic physics are followed to their logical conclusions, irrespective of whether the contemporary conceptual basis underpinning biology would support them. Today the question “How biology works” is considered a “hard problem” scientifically because it appears very complicated, messy and diverse in a molecular (proteins and ribonucleic acids) size range, the mesosphere, which has been relatively little studied [3]. Instead practical progress in manipulating the process of life made over the past half century is largely attributable to recombinant DNA technology. Therefore molecular biological and genetic research is contingent on the assumptions inherent in the gene-focused Modern Synthesis, or neo-Darwinism, but important questions, including what the nature of life is [4] and how it originated [5], remain unresolved. Instead, the priority is to handle the vast quantities of data being produced by genome sequencing and to relate the data to perceived traits ranging across the board from physical characteristics, through pathological illnesses, to cultural, social and even political, preferences. However, these implicitly deterministic assumptions and undertakings have been questioned [6] and challenged at a foundational level [7, 8].

¹ See:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213705/dh_132382.pdf

The proposal

The predominant metaphor for the biological cell is the man-made machine. It justifies a materialist and reductionist/constructionist approach to biology [9]. While reductionism to identify fundamental factors and causal relationships is legitimate, the constructionist phase is problematic because of the phenomenon of emergence. Due to the fact that symmetry² breaking is involved in growth, differentiation, proliferation, etc. in cells and in development of organisms “*in general the relationship between a system and its parts is intellectually a one-way street*” [10]. This is because new qualitative properties, which are by no means easy and usually impossible, to predict, emerge when symmetry is broken. Yet the missing ingredient is the obvious, but often overlooked, photon absorption from the surroundings that invariably leads to the breaking of a system’s symmetry [11]. The incorporation of a single photon may open up entirely new paths for free energy consumption, e.g., as chemical bonds that constitute new compounds which provide new means for the organism to surpass its rivals.

It is self-evident that cells are thermodynamically open to acquire or expel energy, in the form of information and matter from and to their environments. Specifically living systems *dissipate* energy inwardly because they exist in environments that are richer in energy than they are. These dissipative features, the metabolic pathways along which energy is conducted, are termed *actions* in physics. The energy transduction along the paths is governed by the fundamental principle of *least action* for open systems as proposed by De Maupertuis. However, his non-deterministic resolution was subsequently misunderstood and reduced by his successors to a deterministic equation which applies only to stationary systems [11]. In simple terms this principle of Maupertuis requires that any energy gradient will be levelled as efficiently as the prevailing conditions allow. Thus, organisms

in an energy (nutrient) rich environment will transduce as much energy as conditions allow inwards as quickly as possible. This consumption of free energy equates with the 2nd law of maximising entropy [12]. In the case of an open and evolving system competing for free energy with other systems, the most probable state (maximum entropy) will be associated with maximum organisation/complexity when this embodiment minimises the action. Thus, selection will act on variation in free energy consumption, often manifesting itself as high metabolic efficiency, reproduction rate and motility etc. [12]. Since Boltzmann’s time, entropy has been erroneously associated with disorder; however, in thermodynamic terms entropy is a measure of bound and free energy. In a growing organism, just as in an increasing population, entropy comprises both bound and free energy. Eventually, when the maximum entropy state has been attained, all energy is bound in the soma just as is the case when the maximum population density is attained [13]. In an ecosystem bound energy associated with a species is regarded as a source of free energy, e.g., in a form of food for other species. In open systems the influx of energy from the environment to the system (symmetry breaking) is a cause of new properties as is evident by the difference in properties between an un-reacted and a reacted mixture of chemicals [14]. For example, gaseous oxygen and hydrogen, when reacted (exothermically) produce liquid water, or gaseous hydrogen and nitrogen, when reacted (endothermically) produce pungent ammonia, a liquid, or nearly so, at room temperature.

On the basis of the equation for evolving systems, i.e., Maupertuis’ principle [12], it can be concluded that any natural process, from individual organisms through ecosystems, to global biota, can be seen to tend towards increasing complexity due to natural selection favouring the *least action*, the most efficient free energy consumption [12]. Thus, organisms are no ends in themselves, but seen to be pre-eminently the manifestation of the supreme law of physics.

The thermodynamic tenet proposed here should not only be regarded as an intriguing option, but reasoning that is fully consistent

² Symmetry in this context is invariance of the system when viewed from different perspectives. Liquid water, at the molecular level, looks the same from every geometrical perspective, but ice does not; the symmetry has been broken by the phase transition.

with observations. Specifically, the evolutionary equation reproduces ubiquitous patterns, i.e., skewed distributions, logarithmic spirals, sigmoid curves, branching structures, scale-free networks and power laws that make no distinction between animate and inanimate objects. Therefore, we argue that the 2nd law thus is demonstrably a secure foundation upon which to re-build the science of living systems.

Relevant implications

The implications that follow from this insight into physics are profound for biology. The emphasis on the acquisition of free energy and, therefore, metabolism, as the driving force of evolution implies that the origin of life was “metabolism-first”, rather than “replication-first” [15]. According to the 2nd law replication is no objective itself, but a means to consume more free energy more rapidly than, for instance, by simply growing bigger [16] (see below). A metabolism-first origin of life was first proposed by Oparin in 1926 and further developed by Dyson [17], however without explicitly identifying free energy as the driving force. More recently the abiogenesis has been treated within the framework of complex dissipative systems [18].

In evolutionary terms regulation of energy transduction (metabolism) would be a clear strategy for minimising *actions* within the cell and given a “metabolism-first” origin such regulation would most likely be based on the components responsible for metabolism, namely the gene products: in modern cells, proteins, not genomic DNA. This implies an *epigenetic*³ mechanism for the regulation of the cell and a model for such has been proposed for mammalian cells [19] (applies to any eukaryotic cell) and compared to competing genetic regulatory mechanisms [7].

True replication, in contrast to the simple splitting of a cell into two, is favoured when it allows more effective consumption of free energy. Moreover, the natural selection of least time free energy consumption will favour any improvements in transduction

³ In this context the term epigenetic means “over and above genetics” and has no connection to the more recent meaning attached to the term namely, chromatin and DNA marking.

efficiency that the mere replication might bring about. In other words, in abundant circumstances, evolution tends to progress to greater complexity/organisation⁴ to find more effective means to consume free energy.

Epigenetic cell regulation is based on a minimum free energy attractor state⁵ (representing cellular phenotype) achieved through interactions, according to “rules of engagement”, between active gene products, i.e., proteins derived from peptides by folding and other post translational processes, such as phosphorylation [19]. It is useful to envisage the cell/system as comprising a state space with a dimension for each active gene product arbitrarily calibrated from zero to the maximum expressible activity. The attractor location (a profile of typically up to a few thousand gene products and their activities) in the state space represents the phenotype and the basin of attraction provides robustness to

⁴ There are of course examples of loss of, for example, sight when organisms adapt to living in darkness and certain bacteria have lost complexity when adapting to a specific nutrient rich environments, as in the case of *Mycoplasma pneumoniae* in the lung, but these can also be seen as minimising actions.

⁵ The term attractor is often used in a casual way, but for its use in this context it has been strictly formalised [19] and related to thermodynamic imperative to consume free energy in least time. Moreover, we wish to distinguish from the common but erroneous consent that an attractor would be a predetermined state by arguing that during the process of free energy consumption also the attractor will move from its initial position in the free energy landscape. For example, when a stem cell begins to differentiate due to signals from its surroundings, also the surrounding cells will adapt to changes in the differentiating cell. In other words, the non-determinism in the free energy consumption follows from the fact that everything depends on everything else. We work this valuable insight to the powerful notion of an attractor. While it is essentially a *process* of gene product interaction, it can be thought of as the profile of active gene products (proteins), at any point in time, which is regulating the cell and which, because it is inherited at cell division, is also engaged in the inheritance of cellular phenotype. It is postulated that the information on the contributing active proteins is in the form of “rules of engagement” specifying interactions with other proteins, in much the same way the information on enzymes specifies the substrate. As a result, the profile changes with time leading to the evolution of phenotypic properties of the cell. Thus, cell regulation is postulated to be epigenetic rather than genetic [7] and what is inherited at cell division and cell fusion is a *process* in addition to *material*, i.e., DNA. Of course, that dividing cells undergoing differentiation maintain their pre-division state requires that the process be inherited. A property of such a profile is that it is a free energy minimum in the gene product activity state space and, thus, is surrounded by a basin of attraction which endows quasi stability and leads to stable states (attractors) existing only at discrete points in the state space, that is, phenotype is not a continuum [7] and phenotypic/attractor transitions are not gradual.

perturbation of the phenotype. Violation of the rules of engagement between gene products can cause an irreversible attractor/phenotype transition to a variant attractor and, thus, a variant phenotype. Such transitions are discontinuous, i.e., jumps from one thermodynamic steady state to another. The attractor of a cell from a stably replicating species is termed the “home” attractor: it has been evolutionarily conditioned to optimise the integrity of replication [19]. The state space specified by the human genotype potentially contains a very large number of attractors to consume free energy in various circumstances [7].

Although both the DNA and the attractor state are inherited at cell division and fusion [7], the attractor state is the more fundamental of the two in terms of the inheritance of phenotype. This is clear from the fact that in a generic sense there is no contiguous information flow from the genotype to the phenotype [20]: the, in principle non-determinate, peptide folding process [21] acts as an insurmountable barrier to the upward flow of information, i.e., from genotype to phenotype. While genomic sequence stipulates the peptide, subsequent processes under the regulation of the attractor yield the phenotype: in other words the phenotype is its own cause and causation acts downwards on the genome and its products [22]. In terms of physics, the evolution of a system affects the driving forces, which in turn affect the path of evolution. Therefore, the equation of evolution cannot be solved and hence deterministic causation remains an illusion. In fact, when everything depends on everything else it is appropriate to speak about an energy transduction network that evolves as energy flows.

Both the peptide folding and the interaction processes constituting the attractor are dissipative. Therefore, the components of the system denoted as *actions* keep changing due to the influx of quanta, and hence new characteristics will *emerge* [14] in the form of phenotypic properties. Thus, this description that accounts also for the invariable influx would predict that at least a proportion, those involving the interaction of two or more proteins, of phenotypic

properties are *emergent*, i.e., they are not reducible to the properties of the products of individual gene coding sequences (genes) or, indeed, the proteins they give rise to. This is not deny that damage to sequences specifying peptides may lead to damaging phenotypic consequences: simply that the genome contains a proportion of coding sequences that cannot be related to specific phenotypic properties as is expected of a gene as it was originally perceived by Mendel. Moreover, it is worth emphasising that there is no way of knowing *a priori* whether a particular mutation affecting the phenotype will, in some way, be beneficial or harmful for free energy transduction efficiency in a given environment, because the forces imposed by the surroundings cannot be experienced in the absence of energy conduction.

Discussion

The thermodynamic openness of organisms is universally accepted, but curiously it was not emphasized by Darwin. Many of its implications for cellular processing have been and are still, ignored by mainstream cell and molecular biology. Most striking is the assumption that maximum entropy implies maximum disorder as the most probable state. This is the result of Boltzmann’s molecular interpretation of entropy more than 100 years ago. Boltzmann treated his molecular ensembles as systems closed to energy gain or loss with their surroundings. As nothing new can emerge in such stationary-state systems only incoherence, namely disorder, will increase due to exchange of quanta with incoherent surroundings [12]. Conversely, the system will become more coherent via exchange of quanta with coherent surroundings. However, order and disorder are not ends in themselves when an open system evolves towards its most probable state by consuming free energy. Complex and orderly machinery is favoured over simplicity when it is a means to allow more effective consumption of free energy. This is empirically demonstrated by Bénard cells [23] as a fulfilment of the principle of least action. In this case, beyond a threshold of energy gradient within a column of liquid uniformly heated at the base, an ordered

form of convection emerges to increase the efficiency of energy transduction through the column. After more than 100 years of belief that high entropy must mean disorder, it is conceptually difficult to accept that the highly organised structures of living organisms are manifestations of the quest of increasing entropy. Yet it is worth recalling that even Boltzmann noted that animals struggle for entropy, not against it. A Darwin contemporary and naturalist, Edward Blyth, however, did make a reference to the role of nutrient (energy) in promoting fitness saying *“among animals which procure their food by means of their agility, strength, or delicacy of sense, the one best organised must always obtain the greatest quantity; and must, therefore, become physically the strongest, and be thus enabled, by routing its opponents, to transmit its superior qualities to a greater number of offspring.”* This statement can be regarded as recognising the role of the law of least action and could, in fact, be an abstract for the model presented here. Blyth’s comments, although made several years before Darwin published the *“Origin of Species”*, have been dismissed because they implied that natural selection was directed to stabilising species and not to evolutionary change [24]. However, if evolution is regarded as a non-gradual process, (as in, for example, the theory of punctuated equilibrium [25] which is consistent with the thermodynamic account given here) where stress on cellular processes can trigger an attractor transition to a genomically unstable state [19] that may ultimately lead to a new species [7], natural selection can be seen as having both the role to stabilise species (during the “equilibrium” periods, or maintaining the home attractor) and of selecting better adapted organisms in the “punctuations”, or phase of genomic instability. Blyth it seems, like Wallace, got rather less recognition than he deserved in the context of evolutionary theory.

One of the implications of thermodynamic openness that often is ignored is for peptide folding to proteins. In the environment of an energy dissipating system, i.e., a cell, peptides are not bound to fold to the lowest energy tertiary structure present in plain water. Neither are they, contrary to Anfinsen’s dogma [26], constrained to

folding to a structure dictated by the amino acid sequence alone. In an open system protein folding is not a random, i.e., indeterminate, but a non-determinate dissipative process [21], which in the cell is commonly overseen by chaperone and co-chaperone proteins [27]. In general, therefore, no predictions of tertiary protein structure are possible from the information contained in the DNA sequence alone. Furthermore, recent evidence shows that there is far from a one to one agreement between the transcriptome and the proteome [28].

The tenet presented here indicates two inter-related reasons why it has so far not proved possible to forge a clear relationship between genotype and phenotype, except in the case of a limited number of coding sequences. In these latter cases it appears a sequence uniquely specifies a peptide that folds to a single specific protein and that protein acts alone in the phenotype. An example is the rare disease Ehlers-Danlos syndrome, where a mutated peptide is unable to fold into a protein with the properties (information) necessary to form sound collagen tissue. Mendel’s experiments with pea plants and much of 20th century experimental genetics studying very marked and, thus, easily measured traits, probably fall in the same category and are exceptions to the general rule [29]. The majority of coding sequences in the human genome lead to more than one peptide per sequence through diverse splicing of exons and each of these peptides may, through multiple folding opportunities, lead to more than one protein, which may be activated in a number of ways, for example, by phosphorylation. These proteins interact with each other according to the rules of engagement (information acquired upon folding and post translational processes) to contribute to the output of the attractor [19]. These energy dissipative processes are symmetry breaking and potentially give rise to emergent and irreducible (to the originating DNA sequences and indeed proteins) phenotypic properties.

What is termed the “missing heritability” is manifested as a failure to be able to account for the genetic variation of complex traits (including common diseases) in terms of abnormal alleles containing single nucleotide

polymorphisms (SNPs) as detected through genome wide association studies (GWAS) [30]. GWAS might be described as a shortcut in attempting to relate the genotype to the phenotype without sequencing the whole genome. The failure might, therefore, be due to problems with GWAS as a technique or it may be due to there being no relationship to be detected. Empirical evidence strongly suggests the latter. The raw data in a report of a study comprising 50,000 identical or monozygous (MZ) twin pairs [2] indicate that for four cancers known not to have a strong dependence on family history of the disease, the fraction of concordant pairs, f_c , is less than 0.03 and for cancers with a strong family connection (breast and colon cancers), $f_c < 0.1$. MZ twins share identical genotypes and in a high proportion of cases closely similar environments. Therefore, if genetic risk is a major component of overall risk, values of f_c closer to unity would be expected. On the other hand, if the genetic risk were a small component of overall risk, the co-habitation of the twin pairs would likely be the main contributor to concordance. This is illustrated by chronic fatigue ($f_c < 0.26$) which is likely viral in origin, or diseases where domestic environment is known to strongly influence risk, such as CHD ($f_c = 0.25$) through lifestyle and diet, and lung cancer ($f_c < 0.06$) through secondary exposure to tobacco smoke. The evidence, therefore, does not support a strong component of genetic risk, especially for cancers, which account for ~30% morbidity in populations of industrialised countries.

From the above it is clear that the origin of cellular complexity is not exclusively the genomic input, but everything, in particular the surroundings, because everything depends on everything else. The role of surroundings is obvious in dissipative intracellular process, such as polypeptide folding and protein interaction. The free energy input that powers the development of organisms results in the renowned symmetry breaking [10]. Carl Woese, in a criticism of the current reductionist based cell and molecular biology, has already proposed that cellular complexity relies on protein interaction [31]. In effect, since the emergence of multicellular organisms based on eukaryotic cells, the expansion of

information (complexity) output from the phenotype has been due to the greater complexity of active protein interactions within the attractor, rather than primarily the result of adding more components (ultimately the products of diverse coding sequences) to the attractor. This expandable component of cellular processing can be regarded as responsible for the observed diversity of organisms (see below) and account for the fact that markedly morphologically and functionally diverse organisms can have nearly identical genomic sequences. For example, the mouse has nearly as many gene coding sequences as the human, many the same, with a considerable degree of synteny (sequence ordering in the chromosomes), yet is phenotypically quite distinct. The genomic sequence of chimpanzees is even closer to that of humans (99% concordance where sequences are common to both⁶) yet again there are marked phenotypic differences. This phenomenon is a consequence of two factors; a) while both the DNA and the attractor are inherited at cell division and fusion, as noted above, the attractor is the more fundamental in terms of inherited phenotype because it is the origin of phenotypic causality [22], and b) phenotypic output is contingent on the position of the attractor in the state space of active gene products [19]. Thus, it is proposed that mammalian morphological diversity derives primarily from this expandable reservoir of complexity providing information for cell to cell communication and consequent aggregation of cells into complex body and function plans, i.e., organisms. That is, form and function are determined by the position of the attractor in a (for mammals at least) nearly universal conceptual state space based primarily on the same potential reservoir of peptides, which are ultimately derived from a finite set of coding sequences which are to some degree modifiable for adaptive needs [32] by the phenotype itself.

Rönkkö has demonstrated how emergent lifelike properties can be simulated on the basis of the application of rules of interaction between information bearing particles (“atoms”) [33]. This modelling procedure is

⁶ <http://www.nature.com/scitable/knowledge/library/primate-speciation-a-case-study-of-african-96682434>

not constrained to any particular morphology, thus, in the case of multicellular organisms, where the constituent cellular phenotypes would be the information bearing particles, there would be no constraint on the body plans that could emerge: natural selection, however, would favour those best able to transduce energy from their specific ecosystem/environment. In Rönkkö's artificial life model functions would be simulable by deploying specifically differentiated cells within the body plan with similar interaction rules. Thus, the phenotype of the organism can be seen as deriving from information in the cellular phenotype in self-similarity with the way cellular phenotype is derived from interacting proteins.

At this point it is necessary to consider the role of the ecosystems within which organisms are embedded. Darwin fully recognised in the "*Origin*" the enduring nature of ecosystems and the relative constancy of the "proportional numbers" of the various organisms populating them, but at the same time he recognised their vulnerability to disturbance. He takes the example of grazed heath land near Farnham in Surrey, England, populated by a few isolated clumps of mature Scots pines. Fencing off a section of the land to keep cattle out led to a rapid outgrowth of sapling Scots pines within the fence. Examination of the surrounding heath land revealed the presence of Scots pines, their growth prevented by the grazing cattle. Here again self-similarity is encountered: the stability of the ecosystem is contingent on the relative constancy of the proportional numbers of the correct, in fact, what Edward Blyth identified as "*best organised in terms of agility, strength, and delicacy of sense*", constituent organisms. However, the ecosystem is important in another respect in the proposal presented here; namely in providing the context within which the morphological and functional features of organisms evolve. The principle of least action dictates that organisms will adopt, within the prevailing constraints, the best way available to extract free energy from the ecosystem. That is interpreted here as saying that the information at the level of cellular phenotype evolves (adapts) to optimise the body plans and functions such that they are most efficient at extracting the energy available in

that specific ecosystem: the ecosystem has a role in the emergence of phenotype at the organism and cellular levels as indeed, organisms play a role in evolution of the ecosystem by creating niches for other species to exploit. In conventional biology genetic variation is proposed to account for adaptation to environment, and speciation; in the model presented here it is the variation in the deployment of proteins contributing to the attractor and the possibility of attractor transitions stimulated by stress from the environment that are responsible for macro-evolution, i.e., speciation [19]. It should be noted that attractor transitions caused by violations of the rules of engagement between gene products are not gradual, but rather are "jumps" in which the participation, including the degree of activity, of several gene products can change in a single transition [19]. On the other hand, attractors become conditioned to the environment by a gradual process of adjusting the position of the attractor in the state space, evolutionary conditioning, to optimise the integrity of replication.

It may, therefore, be assumed that the ancestor common to roundworms and mammals provided an expandable potential, through the available peptides encoded in the genome, for exploitation in terms of morphology and function. *C. elegans* adapted to be able to function in most ecosystems (in soil), whereas, the mammalian branch evolved in a more limited range of more sophisticated ecosystems, and were better able to exploit much more fully the opportunities that could be derived from the expandable range of protein interactions. This should have led to the potentially testable situation where a greater *proportion* of phenotypic traits have a one to one association with coding sequences in *C. elegans* compared to say *H. Sapiens*. The challenge here is to make a quantitative assessment of phenotypic output.

Treating ecosystems in terms of thermodynamics Schneider and Kay [34] argue that "*life is a response to the thermodynamic imperative of dissipating [energy] gradients*". Biological development occurs when new pathways (actions) for degrading energy emerge. The authors propose that as a consequence the more

developed an ecosystem the lower will be the re-irradiated black body temperature (free energy) and cite evidence for this in terms of measurements of surface temperatures of various ecosystems, which show a trend to lower values the more developed the ecosystem. Equally well it is known that the “density” and diversity of life in ecosystems varies with latitude given adequate rainfall (cf. tropical and temperate forest⁷) as do the body masses for organisms with few physical constraints on the size to which they can grow, such as snakes: the largest known fossil snake, *Titanboa*, was found in Columbia close to the equator [35].

Given the traction that genetics has had over modern biology for the past 60 years it is easy to forget the extent of the debate prior to that period and since, over whether the origin of heredity lay in the nucleus or the cytoplasm [36]. The model advanced here prioritises inheritance from the cytoplasm but, most importantly, it is the inheritance of “process” not “material”. Of course, it is impossible to demonstrate that point directly because for the newly formed cells DNA is essential to transcribe the peptides needed for cell function. However, that cells are *regulated* from the cytoplasm has been demonstrated and is exemplified in humans. Early experiments showed that enucleated fibroblasts could survive *in vitro* and appear normal in all respects other than not having a nucleus [37]. Enucleated fibroblasts with functional hypoxanthine phosphoribosyltransferase (HPRT) enzyme activity are able, through the formation of gap junctions and the transfer of nucleotides or their derivatives, to correct HPRT deficient cells, whereas the karyoplasts (nucleus and remnants of cytoplasm) of the HPRT competent cells were not [38]. These results demonstrate clearly that complex communication between cells can be initiated and take place in the absence of “genes”. Furthermore, erythrocytes (red blood cells) are enucleated as they are released into the blood stream, but are still capable, within their two month lifetime, of exhibiting complex phenotypic features such as circadian rhythm [39]. In fact, a largely temperature resilient 24 hour cyclic

7

<http://www.nature.com/scitable/knowledge/library/terrestrial-biomes-13236757>

phosphorylation of one of the three proteins responsible for circadian rhythm in cyanobacteria can be reconstituted *in vitro* with extracted proteins incubated with ATP [40]. Finally, tardigrades, small water dwelling animals, which exhibit extraordinary resilience to environmental stresses, including ionising radiation [41], seem to owe these properties to being eutelic, i.e., their somatic cells do not divide after hatching from the egg. Thus, once the organism is hatched the DNA of the somatic cells is redundant and damage is of no consequence. Heavily irradiated adults are able to lay eggs but they do not hatch. Developing eggs are radiosensitive in the early stages and only acquire resistance to radiation in the final stage of development [42], presumably when cell division is no longer required. Tardigrades can be regarded as an example of an epigenetically regulated multicellular organism which requires undamaged DNA primarily as a template for replication

If the metabolism-first origin of life is assumed and it is fundamental to the reappraisal described here, then peptide sequence must have been encoded on to DNA at the transition point between “nearly-life” and true life as it exists today, that is, ~3.5 billion years ago. As noted above, without true replication, gains in the efficiency of energy transduction from the ecosystem/environment would not be able to accumulate efficiently (evolution would be much slower, or even not occur) so there would be an obvious advantage in adopting true replication. How this was achieved at the molecular level is unclear, but it is not necessarily an insurmountable obstacle as mechanisms for reverse translation (protein to RNA) have been proposed [43-45] and the protein, reverse transcriptase (facilitating the translation of RNA sequence to DNA sequence) exists. Reverse translation requires the amino acids of the peptide to couple with their tRNA base unit counterparts to form the template RNA that can be polymerised to mRNA. Of course, it is not necessarily the case that the chemistry that preceded true life was based on peptides and so other routes must remain a possibility. As noted by Cook [44] once truly replicating cells existed they had no use for reverse translation so the ability would likely have been lost in the 3.5

billion years of evolution that followed. One potential objection to reverse translation as a means of introducing DNA coding in the context of a metabolism-first origin is how useful peptides were coded, while useless ones were not. It might then be speculated that in the “nearly life” phase DNA/RNA coding for a vast diversity of peptides accumulated, constituting a DNA/RNA data base and those sequences that proved useful led to bacteria at the true origin of life. It is notable that the range of peptides deployed by bacteria is hugely greater than that deployed by mammalian cells: the bacteria inhabiting the human gastro-intestinal tract deploy in excess of nine million different peptides [46]. It also seems likely that bacteria operate on a one to one basis between coding sequence and phenotype, but achieve complex traits by cooperation between diverse species/strains [47]. In contrast, eukaryotic cells have achieved much greater multi-cellular complexity with a very much smaller range of peptides exploiting intracellular cooperation. This is not to deny a role for the very different organisation of the DNA and the presence of membranes, organelles etc., in the eukaryotic cell, as, no doubt, the cytosol plays a crucial role in facilitating the protein-protein interactions. It is interesting to note that fossilised communities of cyano-bacteria (stromatolites) and biofilms [48] date back to close to the origin of life so it would seem that some form of multi-cellularity has long been the norm for the life process.

Conclusions

The thermodynamic tenet presented here represents a major departure from conventional thought on the basis for evolution and its products. It refocuses, for purely physical reasons concerning the role of energy in the natural process called life, attention on the role of metabolic processes and, therefore, proteins, rather than DNA, and the cytoplasm, rather than the cell nucleus. First and foremost, the model is based on the physics of dissipative systems, fully embracing the implications at the molecular level of thermodynamic openness and the quest of attaining stationary status of the cell/organism/ecosystem with its surroundings in least time. A supreme law of

physics governs the life process, namely, the law of least action equivalent to the 2nd law and has, through the former harnessing natural selection and the latter being responsible for producing entropy in the form of bound energy (matter), resulted in the diversity of organisms extinct and extant. The result of this evolutionary processes can be viewed on three levels namely, the cell, the organism and the ecosystem, each level, cellular phenotype, organism phenotype and “ecotype” being represented by an attractor state comprising respectively proteins, cells and organisms, yielding through self-similar processes, emergent properties at the higher respective levels.

Growth and reproduction can be seen as processes that in the natural ecosystem context most efficiently dissipate the incident free energy from the Sun, each level of the hierarchy seeking the minimum free energy state in relation to its own environment, i.e. proteins within cells, cells within organisms, *et cetera*. In this context, the emergence of *H. sapiens* from the Stone Age onwards, some six thousand years, 17 thousandths of a percent of the total duration of life on Earth, most probably represents a unique departure from that which prevailed before. Many species contribute their proportional numbers to more than one ecosystem, *C. elegans*, for example, being almost ubiquitous in soils, but has any other species than *H. sapiens* so grossly over contributed its proportional numbers to the extent of severe disturbance and obliteration of some ecosystems and the annihilation of so many other species? For example, a current threat to marine ecosystems is the increasing dominance of jellyfish over other organisms, most likely resulting from over fishing and eutrophication by washed-off nitrogen fertilisers [49]. On the basis of the arguments above, ecosystem endurance has been integral to the evolutionary processes of adaptation that has improved the efficiency of energy transduction and entropy production. Prior to the Stone Age it would seem that climatic change has been the primary threat to the endurance of ecosystems. Each ecosystem has its top predator, but clearly they have not, in general, abused that position or *H. sapiens* would not have evolved.

Within this framework for biology the gene, as it is generally regarded, is a merely mechanistic, not a profound concept. While gene coding sequences are inherited they are not the “units of inheritance” discovered by Mendel: those are the processes that contribute to the attractor which represents, at the cellular level, the phenotype. The emergent nature of phenotype precludes reducing it to the actions of individual proteins. Furthermore, there is no contiguous deterministic pathway between the information in the gene coding sequences, from which the proteins are derived, and the information inherent in the active proteins participating in the attractor. Thus, population genetics is founded on a subset of coding sequences that can be related to phenotype in a statistical sense, but not on the basis of causation or a viable causal mechanism: genetics, as it is understood today, does not have biological significance.

The evolution of complexity is a puzzle of long-standing predicated on the belief that maximum entropy, as dictated by the 2nd law, would mean maximum disorder, as it is ascribed to thermodynamically closed systems⁸. The insight that the 2nd law dictates maximum complexity in open systems capable of evolution goes part of the way to resolving this puzzle. The second part of this solution is the adoption by eukaryotic cell systems of a methodology for achieving more complex outcomes at the cellular phenotype level through complex post translational “protein chemistry”, fuelled by a limited number of peptides and an even more limited number of coding sequences. While the mechanistic details of this “chemistry” are far from unravelled, its existence cannot be doubted on the basis of empirical evidence, logic and the underpinning physics. Perhaps most controversially it leads to the conclusions that cellular function is an emergent property and is expressed through downward efficient causation from the phenotype to the genotype.

⁸ When a system is truly closed, it cannot even exchange energy to become disordered in a disordered environment or ordered in an ordered environment. If, however, the system is allowed to exchange, but not gain or lose energy then the system will attain the same degree of coherence as its surroundings. So disorder is no end in itself, but common because the superior surroundings, i.e., the free space has very little coherence.

Thus, while gene sequencing may assist in understanding the origins of life forms, as antiquarian books help to reconstruct human cultural history, it is predicted here that the health care revolution anticipated by the UK’s Human Genomics Strategy Group [50] of “patient diagnosis and treatment based on information about a person’s entire DNA sequence, or ‘genome’ – becoming part of mainstream healthcare practice” is over-optimistic.

References

1. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010;**141**:210-7.
2. Roberts NJ, Vogelstein JT, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. The predictive capacity of personal genome sequencing. *Sci Transl Med* 2012;**4**:133ra58.
3. Laughlin RB, Pines D, Schmalian J, Stojkovic BP, Wolynes P. The middle way. *Proc Natl Acad Sci U S A* 2000;**97**:32-7.
4. Schrödinger E. *What is life?* Cambridge: Cambridge University Press; 1944.
5. Penzlin H. The riddle of "life," a biologist's critical view. *Naturwissenschaften* 2009;**96**:1-23.
6. Krimsky S, Gruber J. *Genetic explanations : sense and nonsense*. Cambridge, Mass.: Harvard University Press; 2013.
7. Baverstock K. A comparison of two cell regulatory models entailing high dimensional attractors representing phenotype. *Prog Biophys Mol Biol* 2011;**106**:443-9.
8. Noble D. Physiology is rocking the foundations of evolutionary biology. *Exp Physiol* 2013;**98**:1235-43.
9. Nurse P, Hayles J. The cell in an era of systems biology. *Cell* 2011;**144**:850-4.
10. Anderson PW. More is different. *Science* 1972;**177**:393-6.
11. Annala A. All in action. *Entropy* 2010;**12**:2333-2358.
12. Sharma V, Annala A. Natural process-- natural selection. *Biophys Chem* 2007;**127**:123-8.
13. Makela T, Annala A. Natural patterns of energy dispersal. *Phys Life Rev* 2010;**7**:477-98.
14. Pernu TK, Annala A. Natural Emergence. *Complexity* 2012;**17**:44-47.
15. Cech TR. The RNA Worlds in Context. *Cold Spring Harb Perspect Biol* 2012;**4**:a006742.
16. Annala A, Annala E. The significance of sex. *Biosystems* 2012;**110**:156-61.
17. Dyson FJ. *Origins of life*. Rev. ed. Cambridge [England] ; New York: Cambridge University Press; 1999.
18. Annala A, Annala E. Why did life emerge? *Int J Astrobiol* 2008;**7**:293-300.
19. Baverstock K, Rönkkö M. Epigenetic regulation of the mammalian cell. *PloS One* 2008;**3**:e2290.
20. Baverstock K. The role of information in cell regulation. *Prog Biophys Mol Biol* 2013;**111**:141-143.

21. Sharma V, Kaile VRI, Annala A. Protein folding as an evolutionary process. *Physica A* 2009;**388**:851-862.
22. Baverstock K. Life as physics and chemistry: A system view of biology. *Prog Biophys Mol Biol* 2013;**111**:108-115.
23. Nicolis G, Prigogine I. *Exploring complexity : an introduction*. New York: W.H. Freeman; 1989.
24. Schwartz JS. Charles Darwin's Debt to Malthus and Edward Blyth. *J. Hist.Biol.* 1974;**7**:301-318.
25. Gould SJ. *The structure of evolutionary theory*. Cambridge, Mass. ; London: Belknap; 2002.
26. Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973;**181**:223-30.
27. Taipale M, Krykbaeva I, Koeva M, Kayatekin C, Westover KD, Karras GI, et al. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* 2012;**150**:987-1001.
28. Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, et al. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet* 2011;**7**:e1001393.
29. Lewontin RC. *The genetic basis of evolutionary change*. New York,: Columbia University Press; 1974.
30. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747-53.
31. Woese CR. A new biology for a new century. *Microbiol Mol Biol Rev* 2004;**68**:173-86.
32. Shapiro JA. Rethinking the (im)possible in evolution. *Prog Biophys Mol Biol* 2013;**111**:92-6.
33. Ronkko M. An artificial ecosystem: emergent dynamics and lifelike properties. *Artif Life* 2007;**13**:159-87.
34. Schneider ED, Kay JJ. Order from Disorder: the thermodynamics of complexity in biology. In: Murphy MP, O'Neill LAJ, editors. *What is Life? the Next Fifty Years*. Cambridge: Cambridge University Press; 1995. p. 161-174.
35. Head JJ, Bloch JI, Hastings AK, Bourque JR, Cadena EA, Herrera FA, et al. Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. *Nature* 2009;**457**:715-7.
36. Sapp J. *Beyond the gene : cytoplasmic inheritance and the struggle for authority in genetics*. New York: Oxford University Press; 1987.
37. Goldman RD, Pollack R, Hopkins NH. Preservation of normal behavior by enucleated cells in culture. *Proc Natl Acad Sci U S A* 1973;**70**:750-4.
38. Cox RP, Krauss MR, Balis ME, Dancis J. Studies on cell communication with enucleated human fibroblasts. *J Cell Biol* 1976;**71**:693-703.
39. O'Neill JS, Reddy AB. Circadian clocks in human red blood cells. *Nature* 2011;**469**:498-503.
40. Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, et al. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* 2005;**308**:414-5.
41. Horikawa DD, Sakashita T, Katagiri C, Watanabe M, Kikawada T, Nakahara Y, et al. Radiation tolerance in the tardigrade *Milnesium tardigradum*. *Int J Radiat Biol* 2006;**82**:843-8.
42. Jonsson KI, Beltran-Pardo E, Haghordoost S, Wojcic A, Bermudez-Cruz RM, Bernal -Villigas JE, et al. Tolerance to gamma-irradiation in eggs of the tardigrade *Richtersius coronifer* depends on stage of development. *J Limnol* 2013;**In press**.
43. Nashimoto M. The RNA/protein symmetry hypothesis: experimental support for reverse translation of primitive proteins. *J Theor Biol* 2001;**209**:181-7.
44. Cook ND. The case for reverse translation. *J Theor Biol* 1977;**64**:113-35.
45. Mekler LB. Mechanism of biological memory. *Nature* 1967;**215**:481-4.
46. Yang X, Xie L, Li Y, Wei C. More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *PLoS One* 2009;**4**:e6074.
47. Shapiro JA. Bacteria as multicellular organisms. *Sci Am* 1988:82-89.
48. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* 2004;**2**:95-108.
49. Richardson AJ, Bakun A, Hays GC, Gibbons MJ. The jellyfish joyride: causes, consequences and management responses to a more gelatinous future. *Trends Ecol Evol* 2009;**24**:312-22.
50. Human Genomics Strategy Group. Building on our inheritance: genomic technology in healthcare. London: UK Department of Health; 2012.