ONLY DNA? REALLY?

Prologue

For 50 years, since the discovery of the genetic code (information), DNA has dominated mainstream biology and the sequence of bases in the code has been assumed to define biological function through active products, mostly proteins, derived through transcription, from that code. I make the case here that that assumption has been short-sighted and that a good case can be made for the necessary involvement in the translation of genotype into cellular phenotype (biological function) of a second and independent, source of information.

Part I: The deification of DNA

Now that the celebration of Darwin's anniversaries is behind us it seems an appropriate time to take stock of our intellectual understanding of biology. To those familiar with the theory of evolution by natural selection the numerous documentaries and articles about Darwin over the past year provided little that is new. But what was impressive was to be reminded of the incredible diversity of living systems that has evolved over 3.8 billion years. We might ask the question as to what extent our understanding biology can account for this incredible diversity; is our understanding of biology, in the light of the theory of evolution, capable of providing a coherent explanation for what we observe?

To put that question into perspective it is useful to look at another subject where the process of evolution is writ large, namely cosmology and the evolution of the universe. This is the process that has taken place over roughly 13 billion years. Cosmology has produced a theory of the origin of the universe in terms of the big bang theory, for which there is a lot of circumstantial support. We understand the role of a very weak force, gravity, acting upon tiny ripples of material density in the early life of the universe, leading to the aggregation of matter into galaxies containing Suns, which through nuclear fusion reactions have cooked up the chemical elements out of which living systems are made. To be sure questions remain but progress has been impressive.

My thesis is that biology has not made achievements comparable with cosmology. There is no generally accepted theory of the origin of life, no specific theory why life forms should be what they are and most importantly no consensus in detail on how the genotype translates to phenotype. This is not to diminish in any way the progress made by numerous distinguished biologists from the days of Darwin and Mendel onwards. In my view the key question is “are we making the best use of the knowledge we have?”

Darwin's theory of evolution through natural selection we can take as unassailable; the evidence is overwhelming that random variation in life's heritable material can be exploited, by selection, to give rise to increased complexity and the huge diversity of life forms that we observe and also that which must have once lived but has since been extinguished. The problem arises when we move on to attribute that diversity solely to the diversity of information contained in the sequence of bases in the inherited genomic DNA. One of my colleagues, when I suggested that DNA sequence was the sole basis upon which cell and molecular biology was built, told me that no self-respecting biologist would agree to that. There is, for example, the phenomena of chromosome marking and imprinting, and phenomena such as canalisation during development; these are able to modify the relationship between genotype and phenotype. Yes, quite so, but from where does the information that determines the distribution of chromatin marks and the factors that lead to canalisation come ? To the best of my knowledge no one has suggested a source of information in addition to the genomic base sequence.

Since the inspiration for our present understanding of biology is, among others, Darwin, let us look at something he wrote in "*On The Origin of Species”,* chapter III, entitled "*The Struggle for Existence"*. He wrote "*What a struggle between the several kinds of trees must here* [the ancient Indian mounds, in the Southern United States] *have gone on during long centuries, each annually scattering its seeds by the thousand; what war between insect and insect-between insects, snails, and other animals with birds and beasts of prey-all striving to increase, and all feeding on each other or on the trees or their seeds and seedlings, or on the other plants which first clothed the ground and thus checked the growth of the trees! Throw up a handful of feathers, and all must fall to the ground according to definite laws; but how simple is this problem compared to the action and reaction of the innumerable plants and animals which have determined, in the course of centuries, the proportional numbers and kinds of trees now growing on the old Indian ruins!.* He is of course referring to a stable ecology and we may ask from where does the information that enables this ecology to essentially reproduce itself in a stable manner over centuries, come? It is not in the DNA of the contributing species for each species is trying its hardest to propagate its own genes. There is no centre of control that determines the distribution of species; control, if it is present at all, it is diffused or distributed across the whole ecology. What Darwin was describing is the phenomenon of *self organisation*.

Is there a role for self-organisation in the relationship between phenotype and genotype? Up until 2001, with the sequencing of the human genome, the answer would have been definitively no, there was no need for it since genomic DNA sequence deterministically specified cellular function. That at least was the belief and indeed that this would be the case was the justification for the sequencing enterprise that started in 1991. But in 2001 there was a massive collision with reality; there were roughly 4 to 5 times as many known gene products produced by the human genotype as there were genes identified by that sequencing enterprise. Each of the known gene coding sequences was capable

of producing on average 4 to 5 gene products; how this could be done was less of a mystery than how the cell "decides" which of those 4 to 5 products to produce at any given time. Today, chromatin marking is the favoured explanation; genes can be switched on or off by the attachment of methyl and acetyl groups at specific locations on the chromatin. But this explanation has two problems; the first is that so far no source of information in addition to the genomic DNA base sequence has been proposed as the determinant of the marking. The second objection is that chromatin marking, as the regulatory process, operates at the transcriptional stage, but in higher eukaryotic cells the transcripts produced are not active gene products; they need to be translated into peptides, folded into proteins and then activated by, for example, phosphorylation. It is not always the case that the presence of an active gene product in the cell is directly related to its transcription; that cells adapt to their environment in times much shorter than is required to initiate new transcription is a clear indication of post-transcriptional regulatory processes. Chromatin marking does not therefore fundamentally solve the genotype/phenotype dilemma and we need to look elsewhere.

Back in 1949 the physicist Max Delbrück intervened in a discussion of a paper by the geneticist Sonnenborn who had attributed a particular phenomenon to the reproduction of genes that were either favoured or inhibited by environmental factors. Delbrück noted that "*many systems in flux equilibrium are capable of several equilibria under identical conditions. They pass from one stable* [i.e. ordered] *state to another under the influence of transient perturbations*"[1]. Today, for the term "flux equilibrium" we would use "dynamic steady state". Thus, Delbrück was referring the same phenomenon that underlies Darwin's description of the processes that support a stable ecology. However, the most important distinction he was drawing is not immediately obvious. The kind of system to which Delbrück was referring is what is known as *thermodynamically open*, that is, it exchanges matter and energy with its environment whereas the ideas about how genes exert their effects (and Sonneborn knew nothing of DNA at that time) are neutral in that context; the prevailing physics as elaborated by Schrödinger[2] in 1943 had not grasped the implications of thermodynamic openness although Schrödinger recognised there were implications as he describes living systems as "feeding on entropy"..

At the same time another German physicist with strong interests in biology, Ludwig van Bertalanffy, was striving for what he called a *general system theory*,[3] a theory that would cover situations where something that was conceived as an a single entity (an organism, for example), but which was composed of several interdependent or interacting parts or components (cells and tissues, for example), could be understood as a whole. This theory was "general" in the sense that it would apply not only to biology but also to such technological entities such as electricity generating and distribution systems.

By the 1960s physicists had recognised the limitations of their science when it came to ensembles of particles where statistical averaging, as in the case of the gas laws, was not appropriate. JD Bernal[4] in 1957 recognised that the answer to this crisis would be a major development going well beyond physics: "*a new world outlook is being forged*". Bertalanffy had long held this view with regard to biology, especially in the context of "organismic" biology, i.e., the higher levels of organisation of cells in the organism. However, the implications were apparently not universally recognised even a decade later as in 1969 von Bertalanffy says, in a paper on "*General System Theory and Open Systems*" words to the effect that in Germany, as opposed to the USA, there was really no need to point out that living systems are thermodynamically open. He cites the German biologist, Dost, in 1962 saying "our sons already in their premedical examination take account of this matter" meaning the theory of open systems. There is today, strangely even in Germany, no evidence that this obvious fact is recognised as significant in cell and molecular biology.

One possible explanation of why this strange situation exists, this "forgetting of the significant past", is the bedazzlement of the biological community and beyond, by the astonishing persuasiveness of the structure of DNA embodying the genetic code and its semi-conservative replication. DNA has been credited with powers well beyond what it realistically deserves – it has been deified beyond all reason and we can see this if we look at the cell as a system of interacting components (the active gene products) which is thermodynamically open, which it surely is.

Part II: The cell as a thermodynamically open system

Ludwig van Bertalanoffy identifies two kinds of system, namely those that are reliant on feedback and those that are essentially dynamic. A thermostat is a "feedback reliant" system which is open to information but not energy, therefore not thermodynamically open, although homeostatic. There are in biology system aspects that comply to this model, the body temperature control, for example. But thermodynamically open systems differ fundamentally from all thermodynamically closed systems, which in the long run *must* attain a state of thermodynamic equilibrium, compared to open systems that *may* also attain other non-equilibrium steady states, as noted by both Darwin in connection with ecology and Delbrück at the genetics meeting in Paris in 1949. It is argued here that the cell belongs in this category .

Due to the lack of appreciation of the relevance of this kind of system to biology the language (terminology) in which the subject needs to be discussed is generally not well known. Essentially, the phenotype (of a cell) is the *state* (of the system) which is represented by a *pattern* or *profile* of active gene products present in the cell, each within a specific range of activity. Such a stable state is termed an *attractor* because the adjacent states surrounding it drain into it and thus some impetus or perturbation beyond a certain limit is needed to "release" the system, that is, to induce a change in state, i.e., a phenotypic transition. Typically in a human cell the profile consists of some few thousand active gene products out of the ~100,000 available. The attractor is contingent on what are termed rules of engagement (RoE) between the active gene products[5]. These are essentially causal relationships that

determine which gene products will be active and within what ranges and are of the form **"IF at time t_1 the activity of active gene product "a" is within a specific range r_a THEN active gene product "b" will be in the range r_b at time t_2"where t_2 is greater than t_1.** As such they are in fact "information", what is more they are not dependent on the genomic DNA sequence. To see this we have to consider their origin.

This brings us to the question of the origin of life. Many proposals have been made but none can so far be judged more plausible than others on the basis of what we understand today about biology. It seems most unlikely that DNA itself was present in the earliest precursors to modern day cells; it is an easily degradable molecule by hydrolysis – indeed on any model an unlikely candidate for the (misleadingly) so called template of life – but it is certainly central to the life process. We can be sure of that because under many circumstances its base sequence is highly conserved over evolution.

One of the earliest proposals was made by the Russian biologist Alexander Oparin in the 1920s. He proposed that life started in oily droplets suspended in the ocean some 3.5 to 4 billion years ago. His theory has been elaborated by the physicist Freeman Dyson. Dyson[6] enlarges on the basic ideas proposed by Oparin proposing that cells evolved from semi-permeable oily droplets containing an aqueous solution of small molecules with an agent that provided binding sites and catalysis of polymerisation. While monomers would pass freely in and out of the droplets the synthesised polymers would be retained. In this manner Dyson proposes that a matrix of chemical (possibly, but not necessarily, protein) reactions, represented by a chemical state, is setup within the droplets. Long-lived, or quasi-stationary, states will have basins of attraction and where two or more such basins exist in a droplet, separated by a high barrier, transitions between states become a possibility. Statistically rare transitions where a sequence of reactions yields a more complex quasi-stationary state constitute metabolic activity. Through this process the droplets would enlarge and undergo division by simply dividing their chemical contents into two droplets. True replication and genes, seen in cells today, are deemed subsequent developments in a two stage origin of life.

The first stage would have led to proto-cells capable of metabolism and division, by a purely physical mechanism, but not replication. During this phase there would be the opportunity for the initial two-state nature of the proto-cells to increase in complexity to three or more states; there would be no selection as the reserve of nutrient (small molecules in the ocean) would be vastly greater than the matter constituted in the proto-cells. There would be total freedom for the proto-cells to evolve to greater metabolic efficiency and the increased complexity of states they could support. In this increasing complexity lies the origin of true life, which according to Dyson's model would have led first to enzymes and then to coded information to produce enzymes so that true replication could take place. The crux of the argument made here is that these early steady states evolved to multiple steady states, essentially attractors, and it is these that have been inherited (along with the replication template, which today is DNA), from the earliest life forms that almost certainly pre-date DNA.

In today's cells those evolved products of the primitive attractors formed in proto-cells organise, direct or regulate the information coded into the genomic DNA that produces the active gene products, patterns or profiles of which define the cellular phenotype. Cells are regulated epigenetically by attractors that pre-date the earliest true life forms.

Part III: Justification

This is a strong assertion that is probably counter-intuitive to most readers. Those biologists familiar with attractors probable see them as products of networks and not as free standing entities and almost certainly not as receptacles of biologically important information. The first question they are likely to want to ask is how can this assertion be proved, if not as one colleague put it "what kind of a microscope do you need to see an attractor in a cell?" It is worth revisiting cosmology at this point. The creation of the universe is not an experiment that can be run again so the plausibility of the big bang theory rests on the logic of the arguments about what follows from a set of initial conditions assumed to be the case, with the application of the known physical parameters derived from measurements, such as, for example, the red shift which indicates the rate of expansion of the universe, the estimated masses of visible and dark matter and the known force of gravity. Compared to other theories, the steady state theory, for example, the big bang theory is superior. It may be that in the future a new theory will emerge but for the time being the big bang wins hands down on explanatory power and that is what theories of biology will have to mainly rely upon, Karl Popper notwithstanding. Indeed, the physicist and cosmologist David Deutsch notes in his book "*The Fabric of Reality*"[7] that explanatory power of a hypothesis is more important than its predictive power, this being most effectively used to choose between two hypotheses of equal explanatory power.

That said I justify the idea on four grounds, namely:

- Any regulatory process that acts by interacting with the DNA (transcription) cannot be the ultimate regulatory process forming phenotype in higher eukaryotic cells.
- A singe source of information, the genomic DNA sequence, is insufficient to define phenotype.
- Self-organisation in biology in general and in the cell in particular is commonplace.
- Based purely on the kind of information contained in DNA life could not have started itself.

With about 100,000 active gene products it is clear that cells must be regulated otherwise the undeniable order evident in living systems could not be maintained. Ultimately that regulation has to be at the active gene product level if cells are to be responsive on the time-scales observed, namely, minutes. Thus, while regulation at the transcriptional level is *necessary* it is not *sufficient*. The self-organising attractor model provides regulation at the active gene product level and

we even see evidence of that in that many processes that take place in the cell require aggregates of proteins to work at all, for example, the DNA polymerase complex.

To determine phenotype it is my thesis that two independent sources of information are *required*, one is insufficient. Gene products derived from the DNA base sequence can serve one of two purposes, either as factors *defining* the functional status of the cell or, as factors that *regulate* the functional factors. This is the basis for the genetic regulatory network theory. The regulatory factors (proteins) have specific sites that can bind to discrete sequences on the DNA and initiate transcription of downstream sequences. Huang[8] describes this binding and recognition site concept as "hardwiring" of the genome. However, convincing as it may seem it raises the question of what regulates the regulators and then what regulates the regulators of the regulators and so on *ad infinitum*. A single finite source of information (the genomic DNA base sequence) cannot produce a self regulated entity. The self-organising attractor provides the required second source.

We may also legitimately ask why self-organisation has not been invoked in the interpretation of genotype to phenotype when it has been employed as an explanation for morphology, from the work Alan Turing[9] in 1952 onwards. Furthermore, as Karsenti[10] points out many structural aspects of the cell, including microtubule structure are self-organised. The attractor is simply the manifestation of self-organisation and where in contrast to structural aspects of the cell the functional aspect require complex phenotypic transitions.

Few biologists believe that life had a creator or designer; the only alternative is self-organisation. There is no indication that DNA as a molecule is self-organised (just the opposite in fact as hydrolysis and oxidation are constantly tending to degrade the structure and the information it encodes) and indeed, given the way the information (the triplet base code) is "written" into the molecule it is extremely difficult to envisage a way in which that could be self-organised. Attractors are one candidate for a self-organising origin of life much along the lines proposed by Oparin and elaborated by Dyson[6] (although neither used the term attractor that was the essence of their proposal). If such (or some alternative manifestation of self-organisation) were present in the pre-cursor of the living cell it, in some way, must have given rise to the DNA or at least some precursor. If we are to argue that all that is now necessary doe cells to fully function is the DNA, we need a theory as to how all this happened. That so far we don't have. Indeed, as the current model of the cell is based on the machine metaphor and as machines are not self-organised we have to assume that at some stage in evolution the cell was a self-organised entity and the changed into a machine. That, if it occurred, is an event we might have detected.

The above are four solid reasons, underpinned by logic, to question the dogma that living systems are based solely on information contained in the DNA sequence and thus, to invoke another independent source of information. Of course it does not have to be an attractor – there could be other possibilities. There are however certain constraints, not the least having credibility in terms of the origin of life.

Epilogue

I have no idea how many cosmologists there are on this planet but in the past few decades they have made impressive advances in their subject. I have heard that biologists today outnumber all other disciplines put together, so if numbers count, then biology would be expected to be streets ahead of cosmology. A feature of cosmology is the need, if the subject are to be understood, is to engage with challenging ideas such as dark energy, multiple universes, inflating space, etc. (even if you don't accept them as realistic). By comparison cell and molecular biology pose few such challenges. Indeed, the level of physics required by most biologists working does not go far beyond Newton and whole domain called non-linear physics is largely ignored in biology. As the natural world is governed by non-linear physics what passes for physics, as routinely applied in biology, must be a gross oversimplification of reality. That oversimplification obscures the existence of things like attractors although anybody who rides a bicycle participates in an attractor as a part of their daily life.

Indeed, in that one simple and self-evident step of accepting that life is a thermodynamically open process, a threshold is crossed into "another world" of biology that is needlessly invisible, like the dark matter that makes up most of the matter in the universe.

References

1. Delbruck, M. translation of discussion following a paper by Sonneborn, T. M. and Beale, G.H. *Unites Biologiques Douees de Continute Genetique*, 33 - 35 (1949).
2. Schrödinger, E. *What is life?* (Cambridge University Press, Cambridge, 1944).
3. von Berlalanffy, L. *General System Theory: Foundations, Development, Applications (Revised Edition)* (George Braziller, New York, 1969).
4. Bernal, J. D. *Science in History* (Watts, London, 1957).
5. Baverstock, K. & Ronkko, M. Epigenetic regulation of the mammalian cell. *PLoS ONE* **3**, e2290 (2008).
6. Dyson, F. J. *Origins of life* (Cambridge University Press, Cambridge [England] ; New York, 1999).
7. Deutsch, D. *The fabric of reality : the science of parallel universes-- and its implications* (Allen Lane, New York, 1997).
8. H uang, S. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* **31**, 546-560 (2009)
9. Turing, A. M. The chemical basis of morphogenesis. *Philos Trans. R Soc. Lond. B Biol. Sci.* **237**, 37 - 72 (1952).
10. Karsenti, E. Self-organization in cell biology: a brief history. *Nat Rev Mol Cell Biol* **9**, 255-62 (2008).

The Author:
Keith Baverstock PhD is a Docent at the University of Eastern Finland, Kuopio Campus, Kuopio, Finland.

Further information on the ideas presented here can be found in reference 5 above.