# A comparison of two cell regulatory models entailing high dimensional attractors representing phenotype.

Keith Baverstock
Department of Environmental Science
University of Eastern Finland
Kuopio Campus
Kuopio
Finland
keith.baverstock@uef.fi

**Abstract**

Two models for mammalian cell regulation that invoke the concept of cellular phenotype represented by high dimensional dynamic attractor states are compared. In one model the attractors are derived from an experimentally determined genetic regulatory network (GRN) for the cell type. As the state space architecture within which the attractors are embedded is determined by the binding sites on proteins and the recognition sites on DNA the attractors can be described as "hard-wired" in the genome through the genomic DNA sequence. In the second model attractors arising from the interactions between active gene products (mainly proteins) and independent of the genomic sequence, are descended from a pre-cellular state from which life originated. As this model is based on the cell as an open system the attractor acts as the interface between the cell and its environment. Environmental sources of stress can serve to trigger attractor and therefore phenotypic, transitions without entailing genotypic sequence changes.

It is asserted that the evidence from cell and molecular biological research and logic, favours the second model. If correct there are important implications for understanding how environmental factors impact on evolution and may be implicated in hereditary and somatic disease.

Keywords: phenotype represented by attractor; cell regulation; epigenetics; speciation; genomic instability, somatic and hereditary disease.

## 1.0 Introduction

Understanding the relationship between the apparently rigid coding sequence in the genotype and the much more flexible cellular phenotype in higher organisms is an enduringly important problem in biology that is only partly solved. That a single genotype can give rise to more than 200 terminally differentiated cell phenotypes and several times that number of intermediate phenotypes, in the human, is understood to be due to the selective use of gene products but how that selection is achieved is far from clear. Additionally, although these cellular phenotypes are discrete and can generally only make unidirectional transitions spontaneously, that is, for example, from stem to differentiated cell, within a specific lineage they can be reset to a pluripotent state by specific manipulation including nuclear transfer (Takahashi et al., 2007). It is also the case that the environmental conditions, diet, for example, during some stages of development can influence phenotype not only in the developing organism but also for future generations, with implications for human health (Bateson et al., 2004; Gluckman et al., 2009). In other words cells have a degree of plasticity not reflected in the genotype. Furthermore, ionising radiation can induce in the cells of a stable species a phenotypic transition to what is known as genomic instability (Kadhim et al., 1992), a phenotype that appears to be novel, that is, not associated with processes such as normal differentiation. A possibly related phenomenon has been observed in genetically engineered bacterial cells, which can adopt adaptive phenotypes in response to environmental stress (Kashiwagi et al., 2006) without the benefit, because they have been genetically modified, of a "genetic programme".

The question addressed here is "what are the processes that can account for these phenotypic properties?"; in essence, "how is the cell regulated?" and the starting point in answering that question is to understand what exactly is the nature of cellular phenotype. The commonly accepted answer, based on the Central Dogma of molecular biology which postulates information transfer only in one direction, from DNA to protein, proposes that the sum of the properties of these deterministically translated proteins constitutes the cellular phenotype. This proposition raises many questions outside of genetics, such as how the sequences to be translated are specified and, of course, the problem, still unsolved, of predicting a specific folded protein from its peptide amino acid sequence. In 2001, with the completion of the sequencing of the human genome, it became clear that that deterministic aspect of the Central Dogma was even more problematic. The number of gene coding sequences in the genotype was substantially less than the number of functional products they could produce (Carninci, 2008). This further underlines the importance of non-genetic processes vital to the translation of genotype to phenotype.

The realisation that the extent of chromatin marking in eukaryotic cells far exceeded that required for the permanent imprinting of alleles suggested to some that regulation of the second by second transcription of coding sequences, based on such marking, might be the primary mechanism by which phenotypic expression was regulated (Jaenisch and Bird, 2003). However, according to Huang (Huang, 2009) chromatin marking lacks the necessary stability and locus specificity necessary for it to have a regulatory role in gene expression. I draw attention to additional short-comings below.

In 2000 the idea that phenotype could be represented by a self-organised high dimensional attractor state was proposed independently in two publications (Baverstock, 2000; Huang and Ingber, 2000). The idea that attractors might have an important role in biology was not new. Max Delbrück, in his intervention in a discussion on a paper by Sonneborn given at a genetics conference in Paris in 1949, was probably the first to express the concept. Sonneborn 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*" (Delbruck, 1949). However, well before that Darwin, in Chapter III of the Origin of Species, had in effect articulated the same principle in terms of a stable ecology (Darwin, 1859). The subject prior to 2000 is reviewed by Emlen at al (Emlen et al., 1998). Huang and I proposed that phenotype could be represented by a high dimensional attractor in order to explain very specific features of cell biology. Huang et al were concerned to understand how the fates of neighbouring cells in the developing embryo were determined. Neighbouring cells, possibly in contact, may have very different fates, for example, apoptosis, differentiation or proliferation. It was suggested that these different fates were determined by environmental influences on the cell, for example, soluble growth factors in the extra-cellular matrix, which caused transitions between self-organised attractors in the regulatory

network. My argument was based on an attempt to explain the phenomenon of radiation induced genomic instability.

Genomic instability was uncovered by Munira Kadhim and colleagues at the MRC Radiobiology Unit at Harwell (Kadhim et al., 1992). Explanted bone marrow cells were subjected to low alpha-particle fluences (~1 passage per cell on average) and the survivors plated out singly and grown as clones. Subsequent karyotypic analysis revealed, within a single clone, chromosome aberrations in some cells while others exhibited no damage. As it is expected, on the basis of the prevailing radiobiological dogma, that after the first cell division following irradiation any molecular damage will be replicated in all future generations the only conclusion to be drawn was that damage that was expressed in the later cell divisions was in some way "hidden" in the earlier divisions. The term genomic instability was coined to describe this non-clonal generation of molecular damage. The question I addressed was "what is the inheritance mechanism and the source of the latent molecular damage in the cell progeny?"

Both the above, essentially proposals for the regulation of the cell, have been further developed over the past decade, again independently, and two models (Baverstock and Rönkkö, 2008; Huang, 2009) are now available to compare in the light of the evidence derived from experimental cell and molecular biology research accruing over that period.

## 2.0 Materials and Methods

### 2.1 The models

It is helpful to describe Huang's model first as it is based on the familiar original ideas of Monod and Jacob in 1961 (Monod and Jacob, 1961), namely the concept of the genetic regulatory network (GRN) (Babu et al., 2004). This model is referred to here as the "GRN model". The GRN is seen as "orchestrating" or regulating the process of transcription to produce a *profile* of active gene products, mostly proteins, which can be presented as an attractor. The attractors and other features of the GRN architecture are contingent on the structure of proteins and the target DNA sequences and are therefore "hard-wired" in the genome.

Knowledge of the interactions encoded in the GRN can be derived from experimental data on the transcriptome and the dynamics of the network can be described in terms of ordinary differential equations (ODE), also experimentally derivable. In this way the stable (attractor) states of the system are defined as are the potential transitions between attractors; some, as in the case of differentiation, inherent and others, for example, transitions between lineages, excluded by dynamic gradients and barriers in the state space architecture. Influences external to the cell can cause the attractor make transitions to allowed states in an "all or nothing" manner if the perturbation is sufficiently "strong" to overcome the basin of attraction surrounding the attractor.

The model I advanced, termed here the "independent attractor" (IA) model, is a model for the epigenetic regulation of the mammalian cell (Baverstock and Rönkkö, 2008), where the term "epigenetic" is used (as also does Huang) in the generic sense of "over and above" genetics and not with any implications of chromatin marking. The attractor proposed shares the same basic properties as that proposed by Huang in that it provides a stable phenotype, which with suitable perturbation can be caused to transit to other attractors and thus phenotypes. However, unlike Huang's attractor it is not contingent on the genomic DNA, but is self-organised from the active gene products and inherited, along with the genotype, at every cell division and at cell fusion. Furthermore, the attractor provides, through "rules of engagement" (RoE) between the active gene products, the necessary information to regulate the cell, that is, to determine the stable states (phenotypes) the cell can adopt. Like the genotype the RoE are deemed to exhibit selectable variation.

Two properties of the attractor are of particular significance, namely stability (defined as the ability to replicate the genotype with integrity) and robustness (resilience to perturbation). In a stable species it is assumed that these properties have been optimised through evolutionary conditioning (Baverstock, 2000). To emphasise the uniqueness of the attractors representing the phenotypes of cells of established species they have been termed "home" attractors and unconditioned attractors adopted as a result of stochastic perturbation of attractors are termed "variant" attractors. Thus, genomic instability can be seen as the transition from a home to a variant attractor with the concomitant mutator phenotype (due to less than optimum stability) and greater propensity to be perturbed to other variant attractors as experimentally demonstrated by Falt et al. (Falt et al., 2003). In this sense the genomically unstable phenotype can be regarded as an "incomplete" phenotype as once it is released from the home attractor it relatively readily migrates between variant attractors.

It is hypothesised that the antecedent of the modern cell attractor pre-dates the first cell and formed spontaneously along the lines initially proposed by Oparin in 1924 and elaborated by Dyson (Dyson, 1999). Oparin proposed that cells evolved from semi-permeable oily droplets suspended in water and 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. This was the first proposal for a metabolism-, rather than a replication-, first mechanism. Dyson proposes that a matrix of chemical reactions, represented by a chemical state, is set-up within the droplets. Long-lived, or quasi-stationary, states would have had 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 yielded a more complex quasi-stationary state constituted 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 modern cells, are deemed subsequent developments in a two stage origin of life. Kauffman (Kauffman, 1993) proposes similar ideas based on "auto-catalytic nets" which again can be viewed as precursors to attractors. The evolution of these proto-cells into primitive cells would have been a process of random drift (i.e. without selection) leading to increasing complexity of the reaction mix in the droplet and eventually to true cells.

Thus, the modern mammalian cell incorporates many products of evolution that allow it to replicate itself to produce stable species that exhibit highly sophisticated phenotypic features but it is argued that central to this process is the complex of dynamic steady states, present before true cells evolved and which evolved into the modern cell attractor. This attractor is a free standing entity and is formally defined (Baverstock and Rönkkö, 2008) in terms of the RoE; relations of the form:

$$m_{gpa}(t1) \in r_{gpa} \Rightarrow m_{gpb}(t2) \in r_{gpb}$$

where **m** represents the activity of the gene products (gpa or gpb) and **r** the range of activity of the gene products (gpa or gpb) and time **t1** < **t2**. Thus, if a gene product is pushed out of range the viability of the attractor is compromised and a transition to a variant attractor will follow. No changes need to have occurred in the genomic DNA thus the process is purely epigenetic. However, the variant attractor will not then be in the optimal position in the state space and, thus, its stability will be reduced and it will be more error prone, that is, have a mutator phenotype. It will also be less resilient to further perturbations as the variant attractor will not have been subject to evolutionary conditioning and thus further migration in the spate space is likely. These are the experimentally observed properties of the genomically unstable phenotype (Falt et al., 2003). Furthermore, although many functions of the cell may be retained some might be lost and/or others gained in a single step transition, i.e., in a non-gradual process.

## 3.0 Results

### 3.1 Discriminating between the two models

Since 2000, when the two proposals were initially advanced, evidence has accrued which enables a discrimination to be made between them. In part this evidence derives from a specific experiment and in part from a general understanding of the nature of the processes in the cell at the molecular level. Additionally, the GRN model raises an important theoretical consideration.

#### 3.1.1 *The experiment.*

Kashiwagi et al (Kashiwagi et al., 2006) performed an ingenious experiment with bacteria to determine an organism's response to an environmental challenge unanticipated by the cellular genetic programme. A plasmid was inserted into *E. coli* comprising two mutually inhibitory operons containing genes, the enzyme products of which could compensate for two specific deficiencies that could be introduced into the nutrient. Each operon contained a promoter and reporter for their respective genes. Under normal nutrient conditions neither operon was strongly expressed and the cell occupied what was called the W attractor. When introduced to a nutrient lacking one of the two nutrients the bacteria, after a period of an hour or two of reduced metabolic activity, started to grow expressing the inserted gene that was capable of compensating for the deficiency. The authors argue that in depleted nutrient the cells are able to extract from the transcriptional noise and adopt, an adaptive attractor that enables growth by utilising the appropriate operon. This experiment demonstrates that the *E. Coli* can adapt even when it does not possess any evolutionarily developed compensatory programme for regulation of its genetic information. The absence of a hidden hardwired regulatory pathway to utilise the appropriate operon was proved by switching the promoters between the operons. In addition the possibility that the response was contingent on only a few cells adapting was

eliminated by careful study of the kinetics of the adaptive process. The authors, in effect, conclude that the cells are able to adopt an attractor that is not "hardwired" in the cellular genome. This is contrary to expectation based on the GRN model.

A definitive feature of the experiment but one not reported in the publication (Kashiwagi et al., 2006) is what happens when cells in one of the two adaptive attractors divide. If the GRN model is correct it would be expected that the cell division would give rise to a daughter cell in the W attractor, as specified by the GRN, which would then adapt to the deficient nutrient, whereas if the IA model is correct the daughter cell would be in the adaptive state on cell division. Thus, flow cytometry of a dividing adaptive cell population with synchronized cell cycle should definitively discriminate between the GRN and AI models.

3.1.2 *Molecular biological considerations*

Over the past decade experimental observations that impact on how the cell is regulated have been reported. They include:

- Significant numbers of proteins in mammalian cells have indeterminate structures yet they function normally including in regulatory processes, and
- Important regulatory events occur after transcription and translation: although the regulation of transcription is necessary it is not sufficient to be the overall regulatory process for the cell.

There is now clear evidence (Dunker et al., 2008; Fink, 2005; Sugase et al., 2007) that a significant proportion of proteins, especially in eukaryotes, are natively disordered or unfolded proteins which adopt a tertiary structure upon binding to target sites. Thus, the impression invoked by the term "hard-wiring" in the GRN model is inappropriate and it would appear that specific regulatory functions may be carried out by a number of candidate proteins.

However, the above applies to the regulation of transcription which other evidence shows cannot provide the ultimate regulatory process dictating the translation of genotype to phenotype. Perhaps the most compelling evidence comes from the behaviour of sperm that once mature and transferred to the female reproductive tract undergo morphological changes to increase their motility (Wu and Chu, 2008). At this stage any pathway from transcription to an active gene product is blocked by the absence of cytoplasm in the sperm and so the regulation of this morphological change cannot be through transcription. However, sperm carry numerous proteins that are essential to successful fusion and development of the zygote (Krawetz, 2005; Wu and Chu, 2008), and these proteins can be seen as sustaining the attractor defining the sperm phenotype in the IA model during the transfer of the parental male genotype to the zygote. Further evidence supporting the need for regulation at the active gene product level comes in terms of the speed of response to DNA damage by ionising radiation, where labelling of the damage sites with proteins (Shiloh, 2003) occurs within minutes of irradiation. The response in terms of modified transcription takes of the order of an hour to become evident (Watson et al., 2004).

There is, thus, strong evidence that the primary regulatory processes act at the active gene product level as predicted in the IA model.

3.1.3 *A theoretical consideration*

Whether or not regulation of transcription is the primary regulatory process it is essential, as regulation at any later stage, including directly through the active gene products, requires the active products to have been transcribed. Thus, transcription must be regulated. The question is "would a single source of information, the DNA coding sequence, be sufficient to fully specify the phenotype in a self-organising and self-fabricating entity such as a cell, as the GRN model proposes?"

In the GRN model a subset of the transcribed products are proteins with regulatory functions, for example, transcription factors, which through specific binding at sites on the DNA cause downstream coding sequences to be transcribed to produce either more regulatory proteins or functional proteins that contribute to phenotype. This raises the question "what regulates the regulatory proteins?", a question that leads to impredicativity, a vicious spiral *ad infinitum* and the conclusion that to fully specify phenotype in such a system would require an infinite length of coding and therefore, of DNA. Put another way, the set of "regulators of regulators" required to define phenotype would be infinite. Clearly this is an untenable situation.

The GRN model implies that the information encoded in the DNA is the equivalent of algorithms and so the question can be rephrased as "is the cell a universal Turing machine?": that is, does the cell treat the DNA as a "tape" with input for the computations performed by the cell, the output of which is phenotype. Shapiro (Shapiro, 2005) notes that since the cell utilises the DNA in a physical sense by interacting with

proteins as well as it being used as an information source Turing's concepts may not be useful. Others, however, do regard the cell as a universal Turing machine (Deutsch, 1997) on the grounds that the Turing principle, namely that "*it is possible to build a universal computer that can perform any computation that any other physical object* [a cell] *can perform*" is a fundamental principle of nature. In this model the function of a gene is to instruct (to be a programme for) the synthesis of a specific protein (functional or regulatory), a "low level" function, but combined with other genes these "*low-level programmes add up through layer upon layer of complex control and feedback to sophisticated high level instructions*" (Deutsch, 1997). These are the instructions that translate genotype into phenotype. The question is "has this transition from low- to high-level functional activity sufficient experimental and theoretical support to be considered valid?": does it invalidate the untenable implication of infinite coding sequence?

Noble (Noble, 2010) points out that organisms are "interaction machines" not Turing machines noting that there are no computers into which DNA could be fed to generate life except living systems. He concludes there is no way to retain the concept of a genetic programme as envisaged by Monod and Jacob and central to the GRN model.

Rosen (Rosen, 1991; Rosen, 2000) raises fundamental objections to living systems being treatable as Turing machines. His arguments against the Church thesis, essentially equivalent to the Turing principle, find an analogy in Gödel's incompleteness theorems. Rosen (Rosen, 2000) argues that these theorems can be generalised to imply that rote or formal processes internalised in a system (such as those carried out by a Turing machine) alone cannot be adequate to define a living system. In other words a purely syntactic system is incomplete and requires a semantic partner; i.e., a living system requires an environment with which it can interact. Furthermore, this problem cannot be overcome by internalising, i.e., importing from the environment, further formal processes as Deutsch seems to suggest; external referents to the system are the only solution. Louie (Louie, 2005) supports Rosen's contention (Rosen, 1991) that a (M,R)-system[1], an absolute requirement for a model of an organism, must be uncomputable.

Without a second source of information the GRN model is essentially a purely syntactic system. In contrast, the IA model regards the formal component as the RoE, which call upon the *information* encoded in the DNA, which is placed in the in the cellular environment (Baverstock and Rönkkö, 2008) and, therefore, not internal to the cell. The genotype is therefore the semantic component. In this way the objections raised by Rosen are ovdercome. This situation is comparable to a natural language where a grammar (the RoE) deploys a vocabulary (the genotype) to produce meaning (phenotype).

There are, of course, other "second" sources of information that might fulfil this requirement. Chromatin marking and nucleosome location are process that could replace the role of the RoE. There are, however, difficulties in addition to those identified by Huang (Huang, 2009) and mentioned above, namely lack of stability and locus specificity. Chromatin marking (Barski et al., 2007) and nucleosome location (Bai and Morozov, 2010) regulate at the transcriptional stage so cannot influence directly post transcriptional processes and thus suffer from the same problem as the GRN model. Furthermore, it is far from clear where the information that locates the marking and the nucleosomes on the chromatin comes from; if it were from the genomic sequence it would not necessarily constitute a second independent source of information. Finally, studies of nucleosome turnover (Deal et al., 2010) show that in active DNA regions nucleosomes are replaced frequently, a process that erases chromatin marking. However, the authors suggest that the regulation of nucleosome turnover itself may perpetuate active or silent gene states but they do not suggest what regulates turnover rates. It is also possible that information is carried not only in the coding sequence of the DNA. More than 90% of the base sequence consists of non-coding DNA including extensive repeat sequences. It is possible that this encodes some as yet un-deciphered information. Shapiro (Shapiro, 2005) suggests, for example, that the repetitive sequences serve to format the genome for multiple information storage and transmission functions, however, this would not appear to

[1] An (M,R)-system is a relational model of a natural system with features that distinguish it as an organism. M stands for metabolism and R for replacement or repair. Such a system consists of metabolic catalysts that can be replaced or repaired without intervention from the outside (they are closed to the efficient cause).

satisfy Rosen's requirement that the additional information is not integral with the syntax.

3.1.4. Summary

Both experimental evidence and theory favour the IA model over the GRN model. The most compelling aspect of the experimental evidence is the ability of the IA model to account for the observed post translational regulation and the ability of the bacteria in the experiment of Kashiwagi et al (Kashiwagi et al., 2006) to adapt to nutrient deficit without a genetic programme. This result is reinforced by recent publications, e.g., (Yus et al., 2009) and (Barrick et al., 2009), reporting evidence at odds with the paradigm based on GRNs and genetic programming as the sole origin of phenotype. From the theoretical perspective there is a strong argument that a second source of information independent of the DNA coding sequence would be required to generate phenotype and so far the IA model appears to be the only model to provide this.

3.2 Potential problems with the IA model

Inherent in the IA model is a key assumption, namely that the interacting gene products coded into the genotype are capable of giving rise to a state space architecture with a large number of attractors, well in excess of the number evident in modern cells, so that novel adaptive or variant attractors are available. Whether or not this condition is met in practice is not immediately obvious.

Determining the number of attractors in large dynamic networks is problematic due to the number of possible states of the system. In a comparatively simple random Boolean network (RBN) with N nodes there are $2^N$ possible states. Kaufmann (Kauffman, 1993) showed that such networks with a connectivity K ~=2 exhibit √N state cycle attractors (where the system becomes "trapped" into a repetitive sequence of states). In the case of the human genome with ~100,000 nodes this would be ~317 state cycle attractors. Each of those attractors contains on average √N states. This led Kaufmann to equate these state cycle attractors with the cell cycles of the more than 200 differentiated human cell phenotypes. If that were valid then the independent attractors discussed here would be analogous to the states that comprise the state cycle attractors. However, more recent work shows that in RBNs the number of state cycle attractors increases at least linearly with N (Bilke and Sjunnesson, 2002; Samuelsson and Troein, 2003). Intuitively, more complex systems, such as the IA model proposes, where there is a continuum of values of **m** to be assigned to each node, are likely to have even more attractors than RBNs, simply on the grounds that the number of accessible states is much larger. It is, therefore, argued that the assumption that the state space architecture contains excess attractors is plausible.

The attractor acts as the epigenetic memory as it represents the state of the cell that is inherited at each cell division and at fusion. This raises the important question of how transcription is regulated to sustain the necessary levels of active gene products. It is proposed that intrinsic to the model must be an element of weak downward causation, that is, the state of the system at the level of the output from the attractor dictates or constrains processes at the lower level of the transcription of the coding sequences. There is no obvious mechanism by which feedback from the active gene product level can modulate transcription. The evidence from molecular biology indicates that there are several factors influencing, on the time scale of cellular lifetime, transcription, including chromatin conformation and nuclear architecture (Cremer et al., 2001; Cremer and Cremer, 2001; Fraser and Bickmore, 2007) and marking of the DNA and chromatin (Qiu, 2006). However, exactly what is postulated to be regulating changes in these features of the cell is far from clear. On an evolutionary time scale the spatial distribution of coding sequences among and on the chromosomes may well be ultimately related to the downward influence of the attractor on the transcription process (See below).

**4.0 Discussion**

While both models invoke attractors to represent phenotype there is little or no conceptual similarity between them. The GRN derived attractor is based on the machine metaphor in that it is derived deterministically from the coding sequence of the genotype and therefore not truly self-organised. The independent attractor on the other hand is truly self-organised and based on an alternative dynamic metaphor, for example, "turbulent flow in a liquid" (Woese, 2004) or a whirlpool.

A strength of the IA model is that it plausibly addresses the origin of the cell in terms of the metabolism-first principle, as originally proposed by Oparin and further developed by Dyson (Dyson, 1999). Of course much remains to be discovered about the detail of the link between the cell precursors and the modern cell types, Bacteria, Archaea and Eukarya, the root of which is postulated to be a loose community of progenotes, cells with rapidly mutating genes

exchanged by lateral transfer rather than vertical inheritance and retaining physical division rather than replication (Woese, 1998; Woese, 2002).

The independent attractor, because it is unconstrained by information derived from the genotype, is able to respond to its environment in novel ways and more flexibly than through sequence mutations alone, as is conventionally assumed to be the case. This is the phenomenon that the Kashiwagi et al experiment illustrates, *albeit* in a deterministic (because of the engineered element) rather than a stochastic context. Barbara McClintock (McClintock, 1984) drew attention to this phenomenon in maize where she noted two categories of response, namely programmed, such as the heat shock response and un-programmed or stochastic, such as the response to x-rays. Programmed responses can be seen in terms of attractor transitions that, like those in differentiation, are facilitated by the state space architecture as a result of evolutionary conditioning – Waddington's "necessary paths" or chreods, for example (Waddington, 1977). Un-programmed responses can be seen as stochastic exploration of the state space by the system and that is what the phenomenon of genomic instability exemplifies.

The attractor also modulates the mutation frequency of the cell which is contingent on its location in the state space. Where a cell is optimised for the integrity of its genotypic replication, i.e., it is deploying the optimal combination of gene products, replication errors will be minimised. Such an optimised or home attractor is assumed to be found in an established species which is able to replicate "true to form" apparently indefinitely. Any variant attractor will occupy a less than optimal position in the state space resulting in relatively error prone replication leading to an increased mutation rate – a mutator phenotype and be more readily perturbed leading to the adoption of further variant attractors (Falt et al., 2003), hence the term "incomplete phenotype".

Thus, under the IA model genomic instability rather than being seen as a curiosity is seen as a fundamental aspect of biology: it potentially plays a role in the evolution that gave rise to modern species and is presumably giving rise to further evolution, as well as being a potential factor in hereditary and somatic disease.

The ability of the environment to interact with the attractor to influence phenotype, that is, by forcing gene product activity, **m,** beyond its appropriate range, **r**, could be seen in a germ cell as a first step in a speciation event. The genomically unstable phenotype would undergo a process of conditioning over many generations, in which the stability and robustness of the attractor would be optimised through selection. Concomitantly the mutator phenotype would generate genetic variation. Together these processes could result in a process similar to that described by Gould (Gould and Eldredge, 1993) as "punctuated equilibrium"; a steady "homing in" on a stable species that persists on time scales long compared to the interspecies transitional stages. The possibility of environmentally induced speciation has been widely discussed. For example, in 1986 West-Eberhard (West-Eberhard, 1986) proposed the "alternative-adaptation hypothesis" of speciation. The idea was stimulated by the observation that there were species that can exhibit two or more markedly different phenotypes within a life-span, although not of course simultaneously, for example, butterflies. Transitions between phenotypes are stimulated by environmental conditions. It was hypothesised that a "covariant character set" could develop within individuals in a population, initially silently but be triggered into expression by some environmentally influenced "switch" mechanism to yield a new species in what would appear to be a single step. The IA model would support that "switch" mechanism in terms of the availability of more than one accessible home attractor in a single organism.

It is clear that environmental conditions, diet for example, during development can have a profound influence on the health of the developing organism and in some cases its offspring (Gluckman et al., 2009). For example, people whose birth weights are at the lower end of the range for humans have higher risks of coronary heart disease and type 2 diabetes (Gluckman and Hanson, 2004). The offspring of mice with a paternal radiation history exhibit evidence of genomic instability in their somatic as well as germ cells ((Barber et al., 2006). A significantly decreased proliferation rate in the embryonic cells of mice with a paternal radiation history has been demonstrated in a model chimera system (Wiley et al., 1997; Wiley et al., 1994). In these cases the instability was inherited by the somatic cells from the environmentally affected germ line; however, there is no reason to assume that instability initiated directly in the cells of the developing organism would not have similar potentially detrimental effects on the adult. Furthermore, Jablonka and Raz (Jablonka and Raz, 2009) cite examples of the transmission of effects from somatic cells to germ cells through abnormal hormone levels and/or the migration between cells of small RNAs. There is, thus, a potential route from environmentally

induced somatic cell defects in the parent leading to modification of the germ cell phenotype and thus to effects in their offspring.

In the IA model phenotypic transitions are not gradual; the inputs to phenotype of several gene products can change in a single attractor transition step. Prior to the modern synthesis in 1942, the idea of gradualism in Darwinian evolution had been challenged by several geneticists, among them Galton and Bateson, but most prominently by Goldschmidt (Goldschmidt, 1982) who maintained that the evolution of species, macroevolution, was not simply an extension of within species adaptation, microevolution, but operated by a separate process that was not gradual. There are parallels to be drawn between the IA model and Goldschmidt's "systemic mutation" concept proposed in 1935, namely that "*a pattern change in the chromosomes, completely independent of gene mutations, nay, even of the concept of the gene, will furnish this new* [that is non-gradual in contrast to the Darwinian view] *method of macroevolution*". The pattern changes involved inversions or translocations of the order of loci on the chromosomes as well as in chromosome number. The result of an un-programmed attractor transition might indeed be ultimately resolved (over many generations) in a "pattern change" in the chromosomes as noted above. Phenotypic change as a result of attractor transitions is certainly independent of gene sequence mutations. In evaluating Goldschmidt's contribution to genetics in the introduction to the 1982 edition of Goldschmidt's book, Gould argues that while systemic mutation as an origin of all new species should be rejected, "Goldschmidt's vision was sound". Gould confirms in his book of 2002 (Gould, 2002) that he regards the systemic mutation hypothesis as false but that Goldschmidt's idea of the influence during development of "rate genes" as contributing to speciation the more important contribution. If the IA model is correct this conclusion may need revision.

An important implication of the IA model is for the future of the cell as the basic building block of life. Traditionally it would be assumed that this depended on the immense number of distinct proteins that can be derived from the some 20 amino acids used in eukaryotic metabolism but the evidence indicates that to date cells have exploited only a tiny fraction of these possibilities (Koonin et al., 2002); in this respect the life process has apparently been very parsimonious. The IA model would indicate that this limited "vocabulary" of gene products can be made to be much more versatile by deploying it with different RoE. If that is so, future evolution will depend upon the availability of novel attractors in a given genotypically derived state space architecture. For RBNs that the number of attractors increases more than linearly with the number of nodes would seem to indicate that in the real world of non-Boolean networks the number of attractors could be huge, even immense, even in the context of existing mammalian cells with the order of $10^5$ nodes.

### 4.1 Conclusions

The IA model provides a single framework to underpin cellular biology embracing cells as we know them today together with their origin as pre-cellular chemical systems. There are many knowledge gaps in this framework but the power of the model lies in its ability to make evident how it is that a rigid and highly conserved coding sequence in DNA, the genotype, can give rise to the phenotypic plasticity and responsiveness to environment that is observed in modern cells.

A definitive test of the IA model in terms of the inheritance of adaptive attractors is possible; however, initially the model can be evaluated in terms of its explanatory power, most notably in understanding:

- the nature of genomic instability and the origin of variability in spontaneous mutation rates between and within species,
- the potential for environmentally induced speciation and,
- the rationale for the theory of punctuated equilibrium.

Additionally, the model has the ability to illuminate aspects of the origins of non-genetic somatic and inherited disease, arising from switches to variant attractors representing phenotypes with abnormal characteristics.

## 5.0 Acknowledgements.

Valuable discussion with Oleg Belyakov, Steve Fromage, Andrei Karotki, Hooshang Nikjoo, Mauno Rönkkö and Mike Thorne are gratefully acknowledged.

## References

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. and Teichmann, S. A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14,** 283-91.

Bai, L. and Morozov, A. V. (2010) Gene regulation by nucleosome positioning. *Trends Genet* **26,** 476-83.

Barber, R. C., Hickenbotham, P., Hatch, T., Kelly, D., Topchiy, N., Almeida, G. M., Jones, G. D., Johnson, G. E., Parry, J. M., Rothkamm, K. and Dubrova, Y. E. (2006) Radiation-induced transgenerational alterations in genome stability and DNA damage. *Oncogene* **25,** 7336-42.

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E. and Kim, J. F. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461,** 1243-7.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823-37.

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R. A., Gluckman, P., Godfrey, K., Kirkwood, T., Lahr, M. M., McNamara, J., Metcalfe, N. B., Monaghan, P., Spencer, H. G. and Sultan, S. E. (2004) Developmental plasticity and human health. *Nature* **430,** 419-21.

Baverstock, K. (2000) Radiation-induced genomic instability: a paradigm-breaking phenomenon and its relevance to environmentally induced cancer. *Mutat Res* **454,** 89-109.

Baverstock, K. and Rönkkö, M. (2008) Epigenetic regulation of the mammalian cell. *PLoS ONE* **3,** e2290.

Bilke, S. and Sjunnesson, F. (2002) Stability of the Kauffman model. *Phys Rev E Stat Nonlin Soft Matter Phys* **65,** 016129.

Carninci, P. (2008) Hunting hidden transcripts. *Nat Methods* **5,** 587-9.

Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C. and Cremer, T. (2001) Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res* **9,** 541-67.

Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2,** 292-301.

Darwin, C. (1859) On the Origin of Species, John Murray, London.

Deal, R. B., Henikoff, J. G. and Henikoff, S. (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328,** 1161-4.

Delbruck, M. (1949) Translation of discussion following a paper by Sonneborn, T. M. and Beale, G.H. *Unites Biologiques Douees de Continute Genetique***,** 33 - 35.

Deutsch, D. (1997) The fabric of reality: the science of parallel universes-- and its implications, Allen Lane, New York.

Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., Vacic, V., Obradovic, Z. and Uversky, V. N. (2008) The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9 Suppl 2,** S1.

Dyson, F. J. (1999) Origins of life, Cambridge University Press, Cambridge [England] ; New York.

Emlen, J. M., Freeman, D. C., Mills, A. and Graham, J. H. (1998) How organisms do the right thing: The attractor hypothesis. *Chaos* **8,** 717-726.

Falt, S., Holmberg, K., Lambert, B. and Wennberg, A. (2003) Long-term global gene expression patterns in irradiated human lymphocytes. *Carcinogenesis* **24,** 1823-1845.

Fink, A. L. (2005) Natively unfolded proteins. *Curr Opin Struct Biol* **15,** 35-41.

Fraser, P. and Bickmore, W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature* **447,** 413-7.

Gluckman, P. D. and Hanson, M. A. (2004) The developmental origins of the metabolic syndrome. *Trends Endocrinol Metab* **15,** 183-7.

Gluckman, P. D., Hanson, M. A., Bateson, P., Beedle, A. S., Law, C. M., Bhutta, Z. A., Anokhin, K. V., Bougneres, P., Chandak, G. R., Dasgupta, P., Smith, G. D., Ellison, P. T., Forrester, T. E., Gilbert, S. F., Jablonka, E., Kaplan, H., Prentice, A. M., Simpson, S. J., Uauy, R. and West-Eberhard, M. J. (2009) Towards a new developmental synthesis: adaptive developmental plasticity and human disease. *Lancet* **373,** 1654-7.

Goldschmidt, R. (1982) The material basis of evolution. In *Silliman milestones in science*, Yale University Press, New Haven.

Gould, S. J. (2002) The structure of evolutionary theory, Belknap Press of Harvard University Press, Cambridge, Mass.

Gould, S. J. and Eldredge, N. (1993) Punctuated equilibrium comes of age. *Nature* **366,** 223-7.

Huang, S. (2009) Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* **31,** 546-560.

Huang, S. and Ingber, D. E. (2000) Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp Cell Res* **261,** 91-103.

Jablonka, E. and Raz, G. (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol* **84,** 131-76.

Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl,** 245-54.

Kadhim, M. A., Macdonald, D. A., Goodhead, D. T., Lorimore, S. A., Marsden, S. J. and Wright, E. G. (1992) Transmission of chromosomal instability after plutonium alpha-particle irradiation. *Nature* **355,** 738-40.

Kashiwagi, A., Urabe, I., Kaneko, K. and Yomo, T. (2006) Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. *PLoS ONE* **1,** e49.

Kauffman, S. A. (1993) The Origins of Order: Self Organisation and Selection in Evolution, Oxford University Press, Oxford.

Koonin, E. V., Wolf, Y. I. and Karev, G. P. (2002) The structure of the protein universe and genome evolution. *Nature* **420,** 218-23.

Krawetz, S. A. (2005) Paternal contribution: new insights and future challenges. *Nat Rev Genet* **6,** 633-42.

Louie, A. H. (2005) Any material realization of the (M,R)-systems must have noncomputable models. *J Integr Neurosci* **4,** 423-36.

McClintock, B. (1984) The significance of responses of the genome to challenge. *Science* **226,** 792-801.

Monod, J. and Jacob, F. (1961) Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* **26,** 389-401.

Noble, D. (2010) Biophysics and systems biology. *Philos Transact A Math Phys Eng Sci* **368,** 1125-39.

Qiu, J. (2006) Epigenetics: unfinished symphony. *Nature* **441,** 143-5.

Rosen, R. (1991) Life Itself: a Comprehensive Inquiry into the Nature, Origin and Fabrication of Life. In *Complexity in Ecological Systems*, Columbia University Press, New York.

Rosen, R. (2000) Essays on Life Itself. In *Complexity in Ecological Systems*, Columbia University Press, New York.

Samuelsson, B. and Troein, C. (2003) Superpolynomial growth in the number of attractors in Kauffman networks. *Phys Rev Lett* **90,** 098701.

Shapiro, J. A. (2005) A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* **345,** 91-100.

Shiloh, Y. (2003) ATM and related protein kinases: safeguarding genome integrity. *Nat Rev Cancer* **3,** 155-168.

Sugase, K., Dyson, H. J. and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447,** 1021-1025.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861-72.

Waddington, C. H. (1977) Tools for thought, Cape, London.

Watson, A., Mata, J., Bahler, J., Carr, A. and Humphrey, T. (2004) Global gene expression responses of fission yeast to ionizing radiation. *Mol Biol Cell* **15,** 851-860.

West-Eberhard, M. J. (1986) Alternative adaptations, speciation, and phylogeny (A Review). *Proc Natl Acad Sci U S A* **83,** 1388-1392.

Wiley, L. M., Baulch, J. E., Raabe, O. G. and Straume, T. (1997) Impaired cell proliferation in mice that persists across at least two generations after paternal irradiation. *Radiat Res* **148,** 145-51.

Wiley, L. M., Van Beek, M. E. and Raabe, O. G. (1994) Embryonic effects transmitted by male mice irradiated with 512 MeV/u 56Fe nuclei. *Radiat Res* **138,** 373-85.

Woese, C. (1998) The universal ancestor. *Proc Natl Acad Sci U S A* **95,** 6854-9.

Woese, C. R. (2002) On the evolution of cells. *Proc Natl Acad Sci U S A* **99,** 8742-7.

Woese, C. R. (2004) A new biology for a new century. *Microbiol Mol Biol Rev* **68,** 173-86.

Wu, T. F. and Chu, D. S. (2008) Sperm chromatin: fertile grounds for proteomic discovery of clinical tools. *Mol Cell Proteomics* **7,** 1876-86.

Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, T., Chen, W. H., Wodke, J. A., Guell, M., Martinez, S., Bourgeois, R., Kuhner, S., Raineri, E., Letunic, I., Kalinina, O. V., Rode, M., Herrmann, R., Gutierrez-Gallego, R., Russell, R. B., Gavin, A. C., Bork, P. and Serrano, L. (2009) Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326,** 1263-8.