International Journal of Astrobiology 2 (2): 131–139 (2003) Printed in the United Kingdom
DOI: 10.1017/S147355040300154X © 2003 Cambridge University Press

Understanding genetic regulatory networks

Stuart Kauffman

The Santa Fe Institute, 1377 Hyde Park Road, Santa Fe N.M. 87505
Department of Cell Biology and Physiology, University of New Mexico Medical School, 915 Camino de Salud, NE, Albuquerque, N.M. 87131, USA
e-mail: stu.kauffman@worldnet.att.net

Abstract: Random Boolean networks (RBM) were introduced about 35 years ago as first crude models of genetic regulatory networks. RBNs are comprised of N on–off genes, connected by a randomly assigned regulatory wiring diagram where each gene has K inputs, and each gene is controlled by a randomly assigned Boolean function. This procedure samples at random from the ensemble of all possible NK Boolean networks. The central ideas are to study the typical, or generic properties of this ensemble, and see 1) whether characteristic differences appear as K and biases in Boolean functions are introducted, and 2) whether a subclass of this ensemble has properties matching real cells.

Such networks behave in an ordered or a chaotic regime, with a phase transition, 'the edge of chaos' beween the two regimes. Networks with continuous variables exhibit the same two regimes. Substantial evidence suggests that real cells are in the ordered regime. A key concept is that of an attractor. This is a reentrant trajectory of states of the network, called a state cycle. The central biological interpretation is that cell types are attractors. A number of properties differentiate the ordered and chaotic regimes. These include the size and number of attractors, the existence in the ordered regime of a percolating 'sea' of genes frozen in the on or off state, with a remainder of isolated twinkling islands of genes, a power law distribution of avalanches of gene activity changes following perturbation to a single gene in the ordered regime versus a similar power law distribution plus a spike of enormous avalanches of gene changes in the chaotic regime, and the existence of branching pathway of 'differentiation' between attractors induced by perturbations in the ordered regime.

Noise is serious issue, since noise disrupts attractors. But numerical evidence suggests that attractors can be made very stable to noise, and meanwhile, metaplasias may be a biological manifestation of noise. As we learn more about the wiring diagram and constraints on rules controlling real genes, we can build refined ensembles reflecting these properties, study the generic properties of the refined ensembles, and hope to gain insight into the dynamics of real cells.

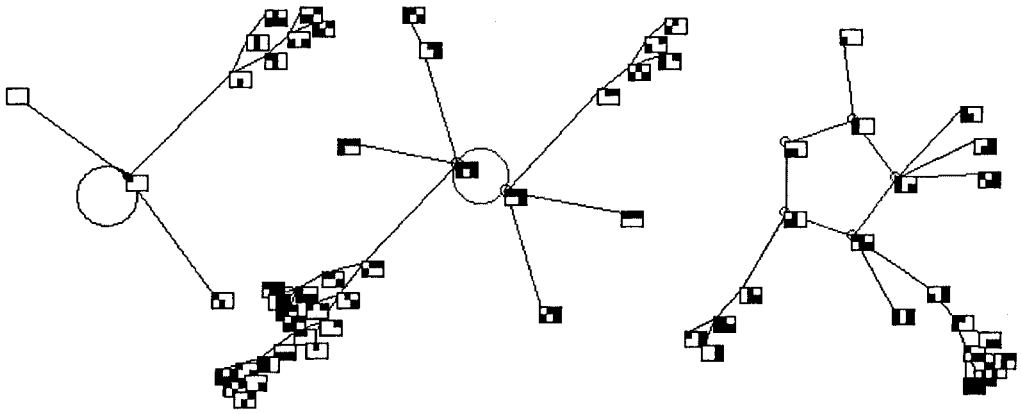
Received 15 June 2003, accepted 30 June 2003

Key words: Random Boolean Networks, ensemble approach, attractors, avalanches, ordered regime, chaotic regime, noise, Derrida plot, frozen component.

Introduction

We are just entering the post-genomic era. The genome project has given us a list of the molecular players at the DNA level. Now we enter upon the daunting task of understanding the integrated dynamical behaviour of the network of genes, transcription factor networks, RNA processing and translational control, and cell signalling cascades among about 40 000 genes and their products. It is fair to say that science has never previously undertaken such a complex task. Consider classical statistical mechanics: one has a volume of gas. All the particles obey the same Newtonian laws. The 19th century triumph of statistical mechanics deeply involved the identity of the laws governing all the particles. But in cells, the dynamical 'laws' or rules, governing the different components of the integrated system differ. We confront a dynamical system the intricacies of which have evolved over more than 3 billion years.

I shall describe and criticize one early approach to understanding such systems, random Boolean networks (RBNs), which I invented almost 35 years ago (Kauffman 1969, 1971, 1974, 1984, 1986, 1993). Such networks were the first to begin to yield potential insights into just the integrated behaviours mentioned above. Work on random Boolean networks is now widespread (Derrida & Pommeau 1986; Derrida & Stauffer 1986; Derrida & Weisbuch 1986; Weisbuch & Stauffer 1987; De Arcangelis & Coniglio 1988; Bastolla & Parisi 1996; Battacharjya & Liang 1996a, b; Socolar & Kauffman 2003) and has been extended to scale-free networks (Serra *et al.* 2003), networks with more than two states per node (Sole *et al.* 2000), networks with piecewise linear dynamical laws (Glass & Hill 1998; Hill *et al.* 1999) and networks with noise (De Arcangelis & Coniglio 1988). See Aldana *et al.* (2002) for a broad review. I will discuss primarily the classical results on random Boolean networks. No one expects real genetic networks to be random. The issues are whether the major



gene	wiring	rule
5	4 0 3	173
4	0 3 1	116
3	4 1 4	114
2	0 2 0	54
1	1 0 4	164
0	5 2 3	234

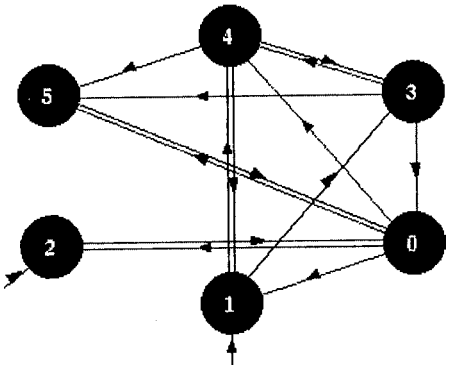


Fig. 1. A graphical depiction of the complete state-space of a sample Boolean net of six genes and $K=3$, assigned at random. There are three basins of attraction, with attractor periods of 1 (fixed point), 2 and 5. The $2^6=64$ states of the six genes are shown within 3×2 rectangles, where active genes are coloured, inactive genes are white. Flows proceed inwards, and then clockwise around attractor cycles. The actual circuitry (wiring) between the six genes (numbered 0 to 5) is shown on the far right, where self-links are short stubs, and in the table on the left, which also shows the logic functions (rules) for each gene, according to Wolfram's convention (Wolfram 1983). The computations and graphics were made with DDLab.

insights gained from the study of RBNs carry over to real genomic systems. It is too early to tell. And it is likely that some aspects will carry over, while others will not. Meanwhile the 'ensemble approach', which consists in sampling random networks from an ensemble of networks built according to whatever constraints we know characterize real genomic systems then analysing the typical, or generic, properties of ensemble members, remains one useful approach to making use of the information we are gathering on real systems to understand their large-scale dynamical and network connectivity implications.

Random Boolean networks

A random Boolean network is a system of N on–off nodes, each representing a gene that can be active or inactive. One global parameter that can be varied is K, the number of regulatory inputs to each gene. K can be constant across all genes, or can be distributed, for example to create a scale-free input distribution. Classically, K is held constant for all N genes. In the simplest case, each gene is assigned at random K inputs from among the N genes, and assigned at random one of the possible Boolean functions of K inputs. These random assignments are then fixed, so that one has sampled at random from the ensemble of all NK networks, with the given values of N and K.

A state of the network is given by the current activities, on or off (1 or 0) of all the N genes. Hence the state space of the network is 2^N. This means that, for the human genome, the possible combinations of gene activities is $2^{40\,000}$. This is about $10^{12\,000}$, a hyperastronomical number. In comparison, the known universe has an estimated 10^{80} particles. The first implication is that real cells must be confined in their dynamical behaviours in this state space to tiny subvolumes. Then a salient question becomes: are there classes of networks that naturally confine dynamical behaviour to such tiny subvolumes. The answer is 'yes'.

Again in the simplest case, time comes in synchronous discrete moments. At each clocked moment, each gene evaluates the activities of its K inputs, consults its Boolean rule, and goes to the activity specified by that Boolean function. Thus, the network, at each clocked moment, passes from

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press

a state of the network to a state of the network. Over time, this typically gives rise to a trajectory in state space. Since the state space is finite, the system must eventually re-enter a state previously encountered. Thereafter, since the system is deterministic, it must persistently traverse a cycle of states, called a state cycle attractor. The number of states on the cycle can range from 1 to 2^N. If released from a different initial state, the system may flow to the first state cycle, or to a distinct state cycle. In general, the state space is partitioned into distinct subsets of states, each subset consists in the basin of attraction of states that either lie on trajectories flowing to a single state cycle or comprise the state cycle itself (see Fig. 1).

Ordered and chaotic regimes

Almost 35 years of research have confirmed that such networks behave in two broad regimes, as a function of K, and biases in the Boolean functions utilized in the network. First, and presumably of deep importance biologically, there is an ordered regime, described shortly. Secondly, there is a chaotic regime, which seems biologically implausible for a variety of reasons discussed below. A phase transition, sometimes dubbed the 'edge of chaos' separates the two regimes. Importantly, the same phase transition between ordered and chaotic regimes has been verified in piecewise linear models as well (Glass & Hill 1998; Hill *et al.* 1999). Thus it seems likely that a vast class of nonlinear dynamical systems all can exist in the ordered regime. My contention is that evolution has made use of this ubiquitous property of integrated dynamical systems to build cells and organisms that rely deeply on the properties of the ordered regime.

A variety of features separate and characterize the ordered and the chaotic regimes. Consider a conceptual movie. Start a network at a random initial state. Over time, if a gene is turning on and off, colour it green. If the gene becomes frozen in the on or in the off state, colour it red. In the ordered regime, at first all genes are green, then an increasing fraction of the genes become red, until a percolating connected red 'sea' spans the network leaving behind green islands of twinkling genes. The size of the red frozen component scales with the size of the network, N.

In the chaotic regime, the same movie reveals a percolating connected green sea of genes that persistently twinkle on and off, and leave behind isolated red frozen islands. The phase transition occurs, as parameters such as K and biases on Boolean rules are tuned, just as the green sea is breaking up into green islands. A conjecture is that the most complex coordinated behaviours can occur just on the ordered side of this phase transition, hence cells or perhaps tissues evolve to the ordered side of the edge of chaos.

Derrida plots

A second way to visualize this phase transition makes use of what I call 'Derrida plots', so named after B. Derrida who introduced them and provided the first theorem showing

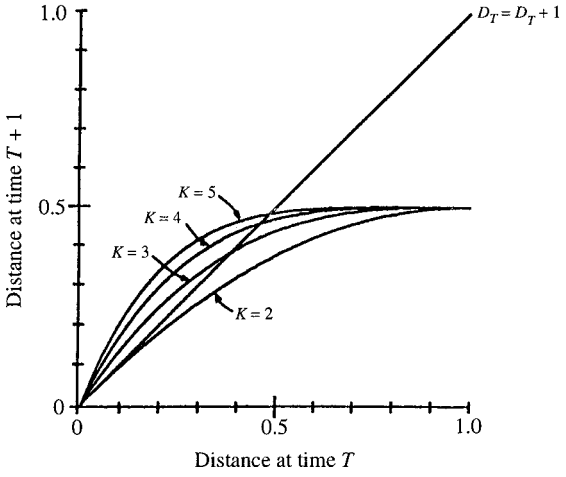


Fig. 2. Recurrence relation showing the expected distance D_{T+1} between two states at time $T+1$ after each is acted upon by the network at time T, as a function of the distance D_T between the two states at time T. Distance is normalized to the fraction of elements in different activity values in the two states being compared. For $K=2$, the recurrence curve is below the 45° line, and hence the distance between arbitrary initial states decreases toward zero over iterations. For $K>2$, states that are initially very close diverge to an asymptotic distance given by the crossing of the corresponding K curve at the 45° line. Thus $K>2$ networks exhibit sensitivity to initial conditions and chaos, not order. Based on the annealed approximation (Kauffman 1993).

the existence of the ordered versus chaotic regimes (Derrida & Pommeau 1986). In effect, a Derrida plot is a discrete analogue of a Lyapunov exponent. Consider two initial states, say (000) and (001) for an $N=3$ network. The Hamming distance between the two states is 1, the number of bits by which the two differ. Dividing by N, the normalized Hamming distance is 1/3. Let each state evolve, via the network, into its successor, say (000)→(100), while (001)→(010). The Hamming distance between the two successor states is 2, and the normalized distance is 2/3. Call the initial distance D_T, and the distance between the successor states D_{T+1}. A Derrida plot, Fig. 2, plots D_T on the X-axis from 0.0 to 1.0, and D_{T+1} on the Y-axis. The pair of initial and successor states given above correspond to a point at $D_T=1/3$, $D_{T+1}=2/3$.

The main diagonal, running from the origin to $D_T=D_{T+1}=1.0$, identifies those pairs of states whose successors are the same distance apart as the initial pair of states. Along this line, there is no divergence or convergence of flow in state space over a single time step.

In the ordered regime, as shown in Fig. 2, if one plots the mean position of points for each choice of D_T from 0.0 to 1.0, one obtains a line that is below the main diagonal everywhere. At the phase transition, the initial slope of the line is 1.0, thus it is tangential to the main diagonal and then curves below it as D_T increases. This means that, on average, states lie on trajectories that are converging in state space. In the chaotic regime, the Derrida curve passes above the main diagonal from small D_T, implying that nearby initial states

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press

diverge apart in the flow in state space. This is parallels the sensitivity to initial conditions in low-dimensional chaos in continuous dynamical systems.

Numerical evidence, but no theorems yet, show that the two criteria, breakup of the green twinkling sea into islands, and tangency of the Derrida plot, coincide.

The character of state cycle attractors differ dramatically in the ordered and chaotic regimes. It has turned out that $K=2$ networks with a random choice of the 16 possible Boolean functions lie exactly on the phase transition, and have state cycles for which the median length scales as $N^{1/2}$ (Kauffman 1969, 1971, 1974, 1984, 1986, 1993; Bastolla & Parisi 1996). Thus a network of 100 000 genes, and $2^{100\,000}$ power states, or $10^{30\,000}$ states, would flow to an attractor with a tiny 318 states. This is an astonishing order. The system, to misuse an analogy, plunges into a tiny black hole of an attractor in state space.

Cell types as attractors

This brings me to the central interpretation that I have made of RBNs, and other nonlinear dynamical models of linked gene activities in some complex network: I interpret a cell type as an attractor of the RBN or other dynamical system. This is an important, and not proven, step. On the one hand, we believe that real cell types are confined patterns of gene activity. Hence it is utterly natural to identify cell types as attractors. On the other hand, real genetic circuits are subject to molecular noise, hence the precise closure of a state cycle is problematic (Aldana *et al*. 2002). I will discuss the implications of noise below. It leads us to consider whether cell types can be attractors, how stable they are to molecular fluctuations, what the biological implications of noise may be, and if a cell type is not an attractor, what might it be?

If cell types are attractors, and multicelled organisms typically have multiple cell types, then RBNs and their cousins better have more than one attractor. For a long time it was thought, based on my own initial results, that the number of attractors in $K=2$ networks scaled as $N^{1/2}$. This allowed a prediction of the number of cell types in an organism as a function of the number of genes it had. And in fact, the number of cell types does appear to scale as a square root function of the total DNA per cell (Fig. 3). This happy result now appears to be incorrect. Recent work has shown that random sampling of initial states undersamples small basins of attraction, and hence undercounts the total number of attractors in $K=2$ RBN. It now appears that the number of attractors in such networks increases somewhat faster than linearly (Socolar & Kauffman 2003). Meanwhile, the number of genes in cells is not known to be linearly proportional to the DNA per cell due to 'junk' DNA.

By tuning the number of inputs, K, and biases on Boolean functions, as described below, it is possible to tune networks to lie at different positions in the ordered or chaotic regimes. Thus, tuning K higher than 2 moves networks into

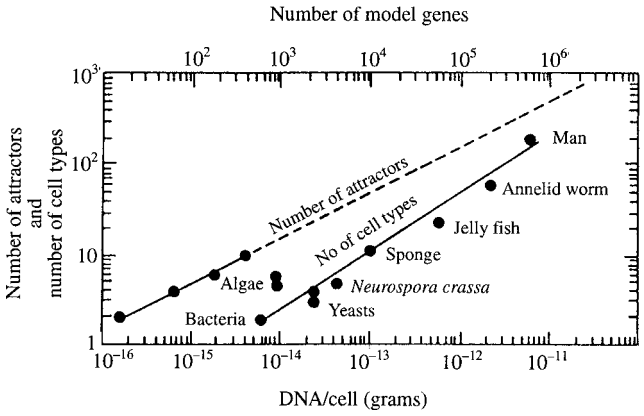


Fig. 3. Logarithm of the number of cell types in organisms across many phyla plotted against the logarithm of the DNA content per cell. The plot is linear with a slope of 0.5, indicating a power-law relation in which the number of cell types increases as the square root of the amount of DNA per cell. If the total number of structural and regulatory genes is assumed to be proportional to the DNA content, then the number of cell types increases as a square-root function of the number of genes. The number of attractors refers to predictions of numbers of model cell types in model genomic regulatory systems having $K=2$ inputs per gene.

the chaotic regime (Fig. 2). Networks with $K=1$ lie deeper in the ordered regime than $K=2$ nets. Biasing Boolean functions can drive networks with $K>2$ inputs into the ordered regime. Deeper in the ordered regime, the number of attractors increases more slowly, although good scaling laws are not yet available. This means that the RBN model probably can be tuned to curve fit the observed distribution of cell types as a function of the number of genes. Such curve fitting would not be entirely pointless, for the resulting networks would make a number of further testable predictions concerning network structure and behaviour described shortly.

One of the class of predictions that can be made using RBN, or any other dynamical model of the integrated genome, is the size distribution of avalanches of gene activity changes following perturbation to the activity of randomly chosen genes. We define a gene as 'damaged' if its behaviour, following a perturbation to the activity of some gene, is ever different from what it would have been without the perturbation. This allows a definition of the size of a damage avalanche, following such a perturbation. In turn this allows a study of the size distribution of avalanches, which turns out to be a power-law distribution with many small avalanches and few large ones in the ordered regime. It appears that there is a finite size effect such that the largest avalanche scales as something like the square root of the number of genes (Harris *et al*. 2003). Thus, for a system with a human-size genome of about 40 000 genes, the largest avalanches should be of the order of several hundred to a thousand or so. This is biologically plausible. In contrast, in the chaotic regime, the distribution of avalanches shows both a power-law distribution

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press

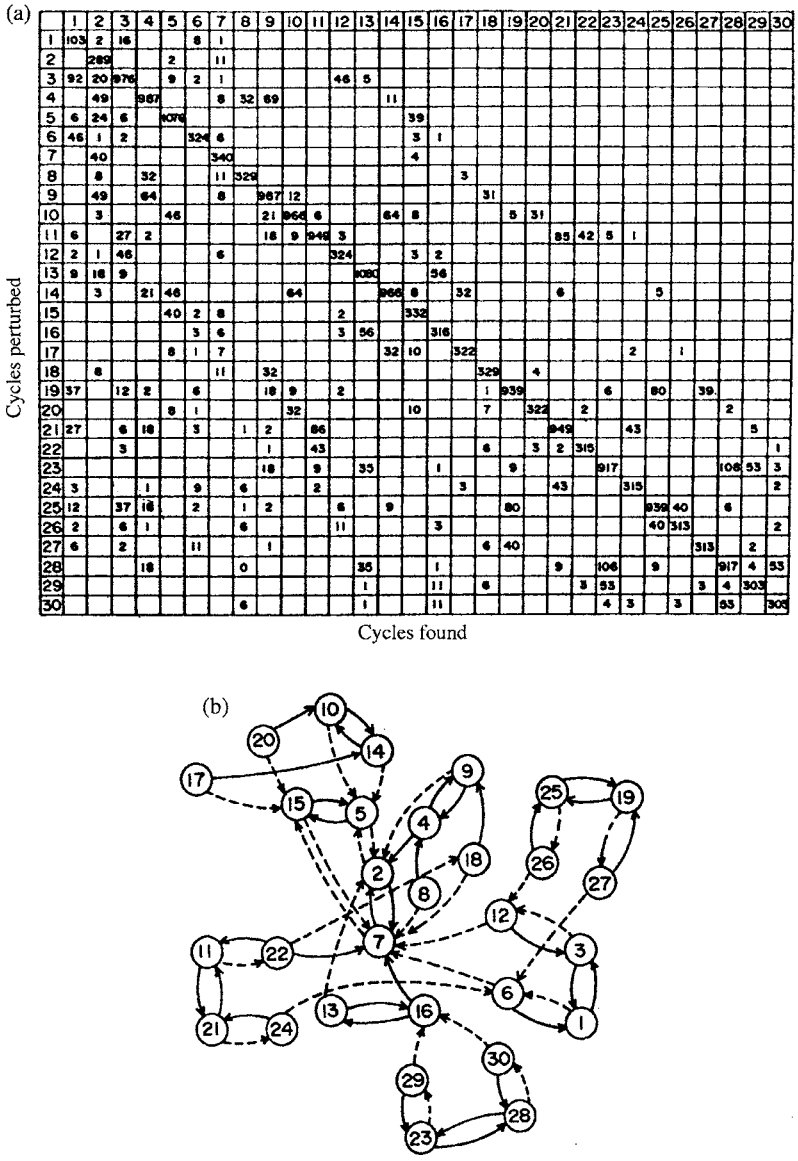


Fig. 4. (a) A matrix listing the 30 state cycles of one network and the total number of times one unit of perturbation, transient reversal of the current activity of a single gene at a single state of state cycle, shifted the network for each cycle to each cycle. The system generally returns to the cycle perturbed and hence exhibits homeostasis. Division of the value in each cell of the matrix by the total of its row yields the matrix of transition probabilities between state cycle modes of behaviour under the drive of occasional random perturbations and constitutes a Markov chain. The transition probabilities between two cycles are often asymmetric. (b) Transitions between cycles shown in (a). The solid arrows are the most probable transition to a cycle other than that perturbed; the dashed arrows are the second most probable. The remaining transitions are not shown. Cycles 2, 7, 5 and 15 form an ergodic set into which the remaining cycles flow. If all transitions between cycles are included, the ergodic set becomes 1, 2, 3, 5, 6, 12, 13, 15. The remainder are transient cycles leading into this single ergodic set. Under the drive of occasional reversal of the activity of any single gene, cell types within the ergodic set can reach one another but cannot reach cell types not in the set (Kauffman 1993).

of modest size avalanches, and a vast spike of huge avalanches were up to 40–50% of the genes change activities. This would imply that modifying the activity of a single gene transiently could trigger alterations in the activities of about 15 000–20 000 other genes. This seems biologically implausible. Meanwhile, such huge avalanches are another signature of sensitivity to initial conditions in the chaotic regime.

Cell differentiation, in any dynamical model that treats attractors as cell types, can consist in at least two kinds of perturbations. In the first, a perturbation moves the system from one attractor to a new basin of attraction from which the system flows to a new attractor. In the second, a perturbation, perhaps arriving from a neighbouring cell, transiently changes the basin of attraction portrait such that the state that the system is on in one cell type in the absence of the

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press

perturbation becomes a member of a new basin of attraction and flows to the new attractor.

Considerable numerical evidence has been studied for the first kind of perturbation in the ordered regime. The typical results are as follows: for most transient reversals of activity of single genes on a state cycle attractor, the system returns to the same state cycle. Thus, cell types show homeostasis to most perturbations as a prediction of the model. For a relatively few perturbations, each state cycle can flow to a few other state cycles. Thus one can achieve differentiation by appropriate perturbations. An important feature of RBNs is that each state cycle can flow to only a few 'neighbouring' state cycles, and from them to some other state cycles (Fig. 4). A directed graph can be drawn showing such flow. It implies something deep: an initial cell type must flow down branching pathways of differentiation to reach a large number of ultimate cell types. But embryogenesis of all metazoans and metaphytons shows exactly such branching pathways of differentiation from the zygote. This raises a deep issue: do organisms follow branching pathways of differentiation because of natural selection, or is this feature of ontogeny so deeply embedded in the self-organized properties of complex gene networks in the ordered regime that evolution is constrained to show this feature? In turn this raises the deep question of whether all the order in organisms is due to natural selection alone, or is it some mixture of natural selection and self-organization. I believe the latter is true, and this requires a revamping of evolutionary theory to include two source of order in organisms, self-organization and selection.

An interesting developmental abnormality, called metaplasia, exhibits the same restricted pathways of aberrant differentiation. Metaplasias are cases where one normal cell type appears in an abnormal location. For example, stomach cells may appear in the esophagus. Fig. 5 summarizes in a directed graph the transformations that occur in humans. An arrow from tissue A to B means that tissue B can metastatically appear in tissue A. Note that each tissue type can transform to only a few other tissue types. Note also that these transformations are not the normal ones of ontogeny. This suggests that the property of having only a few neighbouring tissue types accessible from any one tissue is a deep property of genetic regulatory networks. It seems highly unlikely that natural selection has directly selected for these restricted pathways of metaplasia. Thus either this feature reflects a deep self-organized property of genetic networks, as I hope, or is a side product of selection for some other features of ontogeny.

One plausible picture of ontogeny views different cell types as being characterized by alternative choices of combinations of master genes, in a kind of combinatorial epigenetic code. It is not clear that this hypothesis is correct. Nevertheless, it is interesting that RBN and their continuous cousins, have just the necessary feature to support such combinatorial codes. As noted above, in the ordered regime, a red frozen sea leaves behind isolated green twinkling islands. These are functionally cut off from one another by the red frozen

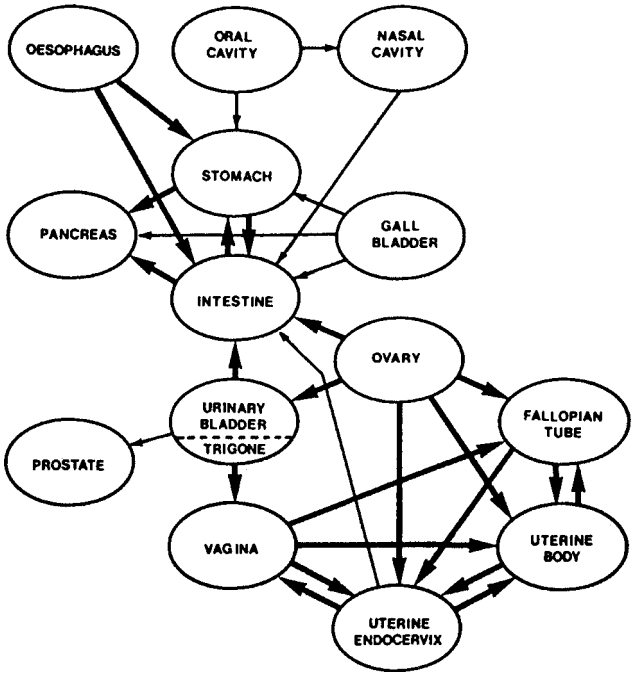


Fig. 5. Graph of homeotic transformations in humans in the epithelial lining of the digestive, urinary and female reproductive systems. An arrow from tissue A to tissue B means that patches of B epithelium can be found in the epithelium of A. Thick arrows denote relatively common events, and thin arrows denote very rare ones. Only the epithelial component of each organ is transformed (Kauffman 1993).

component. Each island is a sub-circuit of the genomic system that has its own alternative attractors. Thus, any state cycle of the entire network is a combinatorial choice of that for one of the attractors of each green island. Thus, the ordered regime predicts that cell types, if they are attractors, are characterized by a combinatorial code of some kind.

Eukaryotic genes appear to be governed by a biased Boolean rule

I mentioned above biases on Boolean functions that can drive the system into the ordered regime in $K>2$ networks. In particular, consider the Boolean OR function of two inputs, A and B, regulating C. The OR functions says C will be active the next moment if A is active now, if B is active now, or if both are active now. Note that if A is active, then C is active at the next moment regardless of the activity of B. I call a Boolean function where at least one input has at least one value which alone can determine the next activity of the regulated locus a 'canalysing' Boolean function. OR is a canalysing Boolean function and both A and B are canalysing inputs, since either alone, by being active, can assure that C is active at the next moment. Good numerical evidence demonstrates that, for $K>2$, networks can be driven from the chaotic to the ordered regime by increasing the fraction of genes that are regulated by canalysing functions with enough canalysing inputs.

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press

A second bias is called P or lambda (Weisbuch & Stauffer 1987). Consider again the OR function of two inputs. Three of the four input combinations are associated with C being active. P is the fraction of the 2^K combinations of input values, for which the dominant output occurs. Thus, 1 is dominant in the OR function, as three of the four cases yield activation of C. P is thus 0.75 for the OR function. For the AND function, P is also 0.75 because three of the four input combinations have 0 as the output. P varies between 0.5 and 1.0. Analytic results show that as K increases above 2, raising P to a critical level leaves the network at the order–chaos phase transition. Further increases leave the network in the ordered regime.

It now appears that real eukaryotic genes are governed far more than at chance levels by canalysing functions (Harris *et al.* 2003). This analysis was carried out for over 150 regulated genes in eukaryotes with $K=3$, 4, 5, and higher numbers of known regulatory inputs. The bias towards canalysing functions is striking, Figs 6(a)–(c), and strongly statistically significant. Genes also show a bias towards high P values. However, an analysis of residual bias towards high canalization after accounting for P, and vice versa, shows that real genes are regulated by canalysing functions tables 1–3. Caveats to this work concern the binarization of continuous data in the papers examined, and that genes regulated by canalysing functions may have phenotypic effects more commonly than genes regulated by non-canalysing functions.

RBN with $K=3$, 4, 5 inputs and random choices of Boolean functions would be in the chaotic regime. In contrast, when we made RBN with the observed distribution of canalysing inputs for $K=3$, 4 and 5 inputs, we found, by a Derrida curve analysis, that the networks were modestly in the ordered regime (see Figs 7a–c). This is evidence that cells are in the ordered regime.

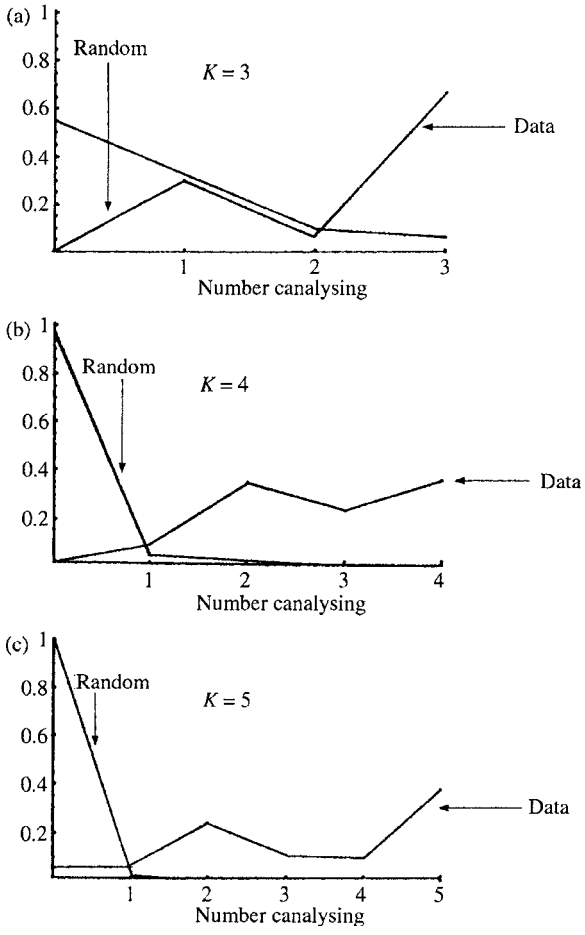


Fig. 6. The upward trending lines are the data, the downward lines represent the distributions of canalysing function from random Boolean functions of K variables.

Critique

More can be said about RBN but it is now time to summarize, criticize and try to imagine how to make further progress. The approach I have taken over the years is an ensemble approach. It is a common approach in statistical physics, for example with spin glasses (Anderson 1985). One constructs members at random from an ensemble of systems and characterizes the generic, typical properties of the systems. The strengths of the ensemble approach are clear. One can derive general features of the integrated behaviour of members of the ensemble that are otherwise not attainable. On the other hand, the approach is statistical. No one can get from an ensemble approach to the statement that gene A regulates gene C in the frog. Thus, an ensemble approach cannot characterize the actual network in real cells. At issue is whether one can refine the ensemble by finding out more about real genetic networks, to the point where some or many of the generic properties of ensemble members characterize real cells. My own biased answer is a reserved 'yes'. There appear to be enough parallels between RBN and their continuous variable homologues to suggest that the ordered regime is readily attainable and exhibits properties likely to have been utilized by natural selection in crafting evolved genetic regulatory systems. On the other hand, my enthusiasm is muted by the realization that 3.8 billion years of crafting can have made networks that are highly atypical of the ensembles one would generate given a set of descriptors of such networks. Only time and further research will tell.

And what of noise? Real genetic systems have relatively few copies of each kind of regulatory molecule per cell (McAdams & Arkin 1999). Hence fluctuations are fundamental to the behaviour of real genetic circuits. For example. S. Leibler (2003) has published results on an experimental 'repressilator', in which three genes repress one another around a cycle of connections. The oscillations of this system are noisy. Noise strikes at the core of the concept that cell types are attractors. In the absence of noise, RBN and deterministic nonlinear dynamical systems typically have attractors and basins of attraction. As noted above, the hypothesis that cell types are attractors is very natural and plausible. But in the presence of noise this pristine picture becomes clouded. For example, for RBN, if one introduced

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press

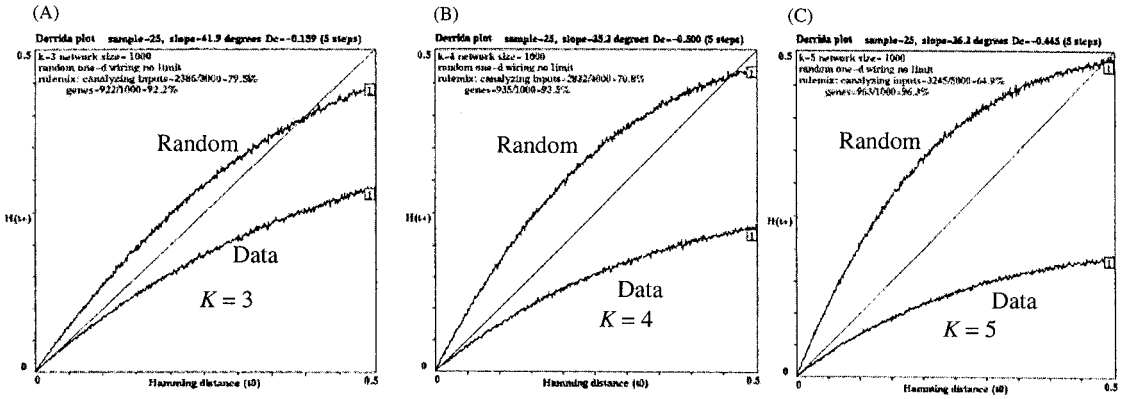


Fig. 7. Comparison Derrida plots using random rules versus rules based on the fraction canalysing rules (f.c.) that were derived from the data for $K=3$, 4 and 5 networks (A, B, C).

a noise term in which each gene disobeys its Boolean rule with probability P, and obeys it with probability $1-P$, then for very small P, one will obtain a Markov chain in which the system jumps occasionally between attractors (De Arcangelis & Coniglio 1988). In this picture, a possible generalization of the concept of an attractor is a set of attractors that can be entered in the Markov process, but cannot be exited (I. Shmulevitch, personal communication). The ease of selecting for such ergodic sets is unclear at this point.

But even granted that, if the noise level is high enough, the concept of an attractor stops having much meaning. At low and intermediate levels of noise the system fluctuates in the vicinity of one attractor, and then undergoes transitions to other attractors. At $P=0.5$ the system wanders randomly in its vast state space.

The existence of noise may not be fatal to the concept that cell types are attractors or noisy attractors with fluctuations. Zhu and colleagues (Zhu *et al.* 2003) have recently analysed the noisy C1–Cro circuit in bacteriophage lambda where each gene represses the other, and shown that that the two steady states, C1 on Cro off, and C1 off Cro on, can be made very stable. For example, they found that with plausible parameter conditions, each state could be stable for a million cell divisions, and have informed me that it is easy to make them stable for a billion cell divisions. Nevertheless, once in a million or billion cell divisions, the cell jumps to the other attractor. Is this biologically reasonable? The answer is currently unclear. Metaplasias, described above, certainly occur. They may be examples of just such noisy transitions. And if one in a million liver cells is a spleen cell, we would have a very hard time finding it. But one is uncomfortable with such a hard to test hypothesis.

On the other hand, it is hard to visualize what besides a noisy attractor a cell type might be. We know that cell types are confined patterns of gene expression. What else might cell types be? At this point the answer is not clear, but warrants serious thought.

With respect to the noise issue, it has been argued that the existence of specific regulatory motifs now being found may function to reduce sensitivity to noise. This raises the terribly difficult issue that real genetic networks are almost certainly non-random in connectivity. For example, in the yeast transcription network, 10 of 106 transcription factors regulate themselves (Lee *et al.* 2002). The mean connectivity among the 106 transcription factors, forming a transcription factor network that drives the system is $K=1$. There appear to be about 103 connections among the 106 transcription factors. Thus, one would not expect 10 self-inputs by chance in a random network. The problem is difficult because, at present, we have only three classes of 'wiring diagrams' well enough defined to study, and none is likely to characterize real genetic nets. We have random nets, scale-free nets and lattices.

It may be that the way forward is to construct networks rich in the kinds of motifs being discovered (Lee *et al.* 2002), tune their relative abundances, fitting overall gene input and output distributions, and study the behaviour of the resulting networks with Boolean, continuous deterministic differential equation, stochastic differential equation and master equation approaches to try to uncover what general wiring diagram rules, such as redundancy, reduce noise and preserve dynamics in the ordered regime.

Summary

Random Boolean networks are of interest as models of disordered dynamical systems and as first crude models of real genomic networks. The existence of the ordered and chaotic regimes appears to be a robust property of RBN and their continuous cousins. Convergent flow in the ordered regime, like error-correcting codes, is a way to cope with the inherent noise in cellular molecular networks. The extent to which this can be successful, and whether occasional metaplastic transformations are inherent in multicelled organisms remain to be seen. The hypothesis that cell types are attractors is natural and plausible, but not free from difficulties nor demonstrated experimentally. Finally, I would note that the predictions of the RBN models, the scale-free cousins and the continuous variable cousins, are now largely testable using gene chip arrays to establish whether cell types are

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press

attractors homeotstatically stable to most perturbations, the size distribution of avalanches, the constrained transitions between constrained patterns of gene activity called cell types and the other predictions discussed above.

References

Aldana, S., Coopersmith, S. & Kadanoff, L.P. (2002). http://arXiv.org/abs/nlin.AO/020406

Anderson, P.W. (1985). Spin glass Hamiltonians: a bridge between biology, statistical mechanics, and computer science. In *Emerging Synthesis in Science: Proc. Founding Workshops of the Santa Fe Institute*, ed. Pines, D. Santa Fe Institute, Santa Fe, NM.

Bastolla, U. & Parisi, G. (1996). Closing probabilities in the Kauffman model: an annealed computation. *Physica D* **98**, 1.

Battacharjya, A. & Liang, S. (1996a). Power-law distributions in some random Boolean networks. *Phys. Rev. Lett.* **77**, 1644.

Battacharjya, A. & Liang, S. (1996b). Median attractor and transients in random Boolean nets. *Physica D* **95**, 29.

De Arcangelis, L. & Coniglio, A. (1988). Critical temperature in Kauffman cellular automata. *Europhys. Lett.* **7**, 113.

Derrida, B. & Pommeau, Y. (1986). Random networks of automata: a simpled annealed approximation. *Europhys. Lett.* **1**, 45.

Derrida, B. & Stauffer, D. (1986). Phase transitions in two-dimensional Kauffman cellular automata. *Europhys. Lett.* **2**, 739.

Derrida, B. & Weisbuch, G. (1986). Evolution of overlaps between configurations in random Boolean neworks. *J. Phyique* **47**, 1297.

Glass, L. & Hill, C. (1998). Ordered and disordered dynamics in random networks. *Europhys. Lett.* **41**, 599.

Harris, S.E., Sawhill, B., Wuensche, A. & Kauffman, S. (2003). A model of transcriptional regulatory networks based on biases in the observed regulation rules. *Complexity* **7**, No. 4.

Hill, C., Sawhill, B., Kauffman, S. & Glass, L. (1999). Transition to chaos in models of genetic networks. In *Statistical Mechanics of Biocomplexity, Proc. XVth Sitges Conf.* I, eds Requera, D., Vilar, J.M.G. & Rubi, J.J., p. 261. Springer-Verlag, Berlin.

Kauffman, S.A. (1969). Metabolic stability and epigenesist in randomly connected nets. *J. Theor. Biol.* **22**, 437.

Kauffman, S.A. (1971). Gene regulation networks: a theory for their global structure and behavior. *Curr. Topics Dev. Biol.* **6**, 145.

Kauffman, S.A. (1974). The large-scale structure and dynamics of gene control circuits: an ensemble approach. *J. Theor. Biol.* **44**, 167.

Kauffman, S.A. (1984). Emergent properties in random complex automata. *Physica D* **10**, 145.

Kauffman, S.A. (1986). Developmental logic and its evolution. *Bioessays* **6**, 82.

Kauffman, S.A. (1993). *Origins of Order: Self Organization and Selection in Evolution*. Oxford University Press, New York.

Lee, T.I. *et al.* (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799.

McAdams, H.H. & Arkin, A. (1999). It's a noisy business. *Trends Genetics* **15**, 65.

Serra, R., Villani, M.K. & Agostini, L. (2003). How the size of the network affects the dynamics of scale-free random Boolean networks, submitted.

Socolar, J. & Kauffman, S.A. (2003). Scaling in ordered and critical random Boolean networks. *Phys. Rev. Lett.* **90**, 6.

Sole, R., Luque, B. & Kauffman, S.A. (2000). Phase transitions in random networks with multiple states. Santa Fe Institute. Working Paper no. 00-02.011.

Weisbuch, G. & Stauffer, D. (1987). Phase transition in cellular random Boolean nets. *J. Physique* **49**, 11.

Zhu, X.M., Yin, L., Hood, L. & Ping, A. (2003). Robustness, stability and efficiency of phage lambda regulatory network: dynamical structure analysis, submitted.

https://doi.org/10.1017/S147355040300154X Published online by Cambridge University Press