

Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets

S. A. KAUFFMAN

*Department of Anatomy, University of California Medical School,
San Francisco, California, U.S.A.*

and

*Research Laboratory of Electronics, Massachusetts Institute of Technology,
Cambridge, Massachusetts, U.S.A.†*

(Received 19 March 1968, and in revised form 8 July 1968)

*"The world is either the effect of cause or
chance. If the latter, it is a world for all
that, that is to say, it is a regular and
beautiful structure."*

Marcus Aurelius

Proto-organisms probably were randomly aggregated nets of chemical reactions. The hypothesis that contemporary organisms are also randomly constructed molecular automata is examined by modeling the gene as a binary (on-off) device and studying the behavior of large, randomly constructed nets of these binary "genes". The results suggest that, if each "gene" is directly affected by two or three other "genes", then such random nets: behave with great order and stability; undergo behavior cycles whose length predicts cell replication time as a function of the number of genes per cell; possess different modes of behavior whose number per net predicts roughly the number of cell types in an organism as a function of its number of genes; and under the stimulus of noise are capable of differentiating directly from any mode of behavior to at most a few other modes of behavior. Cellular differentiation is modeled as a Markov chain among the modes of behavior of a genetic net. The possibility of a general theory of metabolic behavior is suggested.

1. Introduction

A living thing is a complex net of interactions between thousands or millions of chemical species. A fundamental task of biology is to account for the origin and nature of metabolic stability in such systems in terms of the mechanisms which control biosynthesis. In the thermodynamics of gases, the mathematical laws of statistics bridge the gap between a chaos of colliding molecules and the simple order of the gas laws. In biology, a gene specifies a protein, and the

† Present address: Cincinnati General Hospital, Cincinnati, Ohio, U.S.A.

output of one gene can control the rate of output of a second. The mathematical laws which engage large nets of interacting genes into biosynthetic coherence remain to be elucidated.

In this article I report the behavior of large nets of randomly interconnected binary (on-off) "genes". The motives for this choice of model are many.

The analogy of genetic repression and derepression with digital computers has suggested to several authors (Jacob & Monod, 1963; Apter, 1966; Bonner, 1965; Sugita, 1963; Kauffman, 1967) that the genome embodies complex switching circuits which constitute a program for metabolic stability and cell differentiation, rather than providing a coded description of these phenomena.

It is a fundamental question whether metabolic stability and epigenesis require the genetic regulatory circuits to be precisely constructed. Has a fortunate evolutionary history selected only nets of highly ordered circuits which alone insure metabolic stability; or are stability and epigenesis, even in nets of randomly interconnected regulatory circuits, to be expected as the probable consequence of as yet unknown mathematical laws? Are living things more akin to precisely programmed automata selected by evolution, or to randomly assembled automata whose characteristic behavior reflects their unordered construction, no matter how evolution selected the surviving forms?

In this article I present evidence that large, randomly connected feedback nets of binary "genes" behave with stability comparable to that in living things; that these systems undergo short stable cycles in the states of their constituents; that the time course of these behavior cycles parallels and predicts the time required for cell replication in many phyla; that the number of distinguishable modes of behavior of one randomly constructed net predicts with considerable accuracy the number of cell types in an organism which embodies a genetic net of the same size; that, like cells, a randomly connected genetic net is capable of differentiating directly from any one mode of behavior to at most a few of its other modes; and that these restricted transition possibilities between modes of behavior allow us to state a theory of differentiation which deduces the origin, sequence, branching, and cessation of differentiation as the expected behavior of randomly assembled genetic nets.

Mathematical insight into the behavior of randomly connected feedback systems is slight. Goodwin (1963) has treated the gene as a continuously oscillating biochemical element whose output of mRNA is repressed by the protein specified. To study coupled systems of such biochemical oscillators, Goodwin was constrained by the conditions of integrability to restrict cross-coupling between genes to be symmetrically repressive and form a linear sequence in which no gene represses more than its two neighbors. There is no reason, however, to suppose that the crossreactions between real genes are

similarly constrained. To study the behavior of nets with arbitrarily complex couplings requires us to abandon the effort to obtain an integral of motion for the system (Goodwin, 1963).

Several considerations suggest the advantage of modeling the gene as a binary device, able only to be on or off. The most fundamental of measures is the binary category scale. Use of these simplest devices facilitates study of the behavior of truly complex nets; the behavior of randomly connected, but then fixed, nets of binary components should provide a reliable guide to the behavior of similar systems whose components' behavior are described by continuous or probabilistic functions; synthesis of mRNA is, in fact, probably an all or none binary process; the number of repressor molecules per gene is thought to be less than about 12 (Bretscher, 1967), therefore it seems preferable to treat the activity of a gene as a discrete, not continuous, function of its input.

To study the behavior of randomly interconnected nets requires a definition of the population from which equiprobable sampling is to be done. A distinct advantage in the choice of a binary model for gene activity is that the number of different possible rules by which a finite number (K) of inputs may affect the output behavior of a binary element is finite -2^{2K} (see Fig. 1). This

(a)

T	$T+1$
$W \ X \ Y$	Z
0 0 0	0
0 0 1	0
0 1 0	0
0 1 1	1
1 0 0	1
1 0 1	0
1 1 0	1
1 1 1	1

(b)

T	$T+1$
$X \ Y$	Z
0 0	0
0 1	0
1 0	0
1 1	0

1

T	$T+1$
$X \ Y$	Z
0 0	0
0 1	0
1 0	0
1 1	1

2

T	$T+1$
$X \ Y$	Z
0 0	0
0 1	0
1 0	1
1 1	0

3

T	$T+1$
$X \ Y$	Z
0 0	0
0 1	1
1 0	1
1 1	1

15

T	$T+1$
$X \ Y$	Z
0 0	1
0 1	1
1 0	1
1 1	1

16

FIG. 1. (a) W , X , and Y are each binary devices which act as inputs to Z , another binary device. The 3×8 matrix of 1 and 0 below W , X , Y list the eight possible configurations of input values to element Z . The column under Z assigns to it the value it will assume one moment after each input configuration. (a) is one of the $2^{2^3} = 256$ Boolean functions of three variables. (b) The $2^{2^2} = 16$ Boolean functions of two input variables are derived by filling the column under Z with 1 and 0 in all possible (16) ways. Function 1 is *contradiction*, 2 is *and*, 16 is *tautology*.

allows construction of switching nets which are random in two different, but well defined, senses: the K inputs to each binary "gene" may be chosen at random; the effect of those inputs on the recipient element's output behavior may be randomly decided by assigning at random to each element one of the possible 2^{2^K} Boolean functions of its inputs. Once built, the nets I have studied remained fixed in the choice of inputs to each gene, and their effect on its output.

The number of genes whose products directly affect the output of any gene is not known. Therefore, I have studied nets in which each gene has direct inputs from all genes, nets with one input per gene, nets with two inputs per gene, and nets with three inputs per gene.

Since the autonomous, undriven behavior of a system must be elucidated before the effect of exogeneous inputs can be understood, I have studied the behavior of switching nets free of external inputs. A bacterium in a constant environment undergoes autonomous changes in the concentrations of molecular species, and the sea urchin, in a similarly homogeneous surrounding, develops in an orderly sequence of states from its zygote. Since constant external input to a net is equivalent to a similar net held free of external input, stable oscillations of chemical species and cell differentiation seem to be largely autonomous behaviors of metabolic nets.

The study of randomly constructed but deterministic switching nets forms a poorly developed area of automata theory. Walker & Ashby (1965) have examined the effect of the choice of Boolean function on the behavior of randomly interconnected nets of binary elements. They simulated nets in which each of the 100 elements received a feedback input from itself, and randomly assigned inputs from two other elements. For each experiment, all elements of the net were assigned the same Boolean function.

These nets embody behavior cycles (described in detail below). They found that the choice of the Boolean function assigned to all the elements markedly affected the length of these behavior cycles. Some functions (e.g. "and") yield very short cycles, others (e.g. "exclusive or") yield cycles of immense length.

Since there is no reason to suppose that, in living genetic reaction nets, all elements are assigned the same Boolean function, I have studied nets in which all the 2^{2^K} possible Boolean functions are assigned randomly, one to each element.

2. Genetic Model

On these considerations, the gene is modeled as a binary device able to realize any one, but only one, of the possible Boolean functions of its K inputs. If the activity of a formal gene, for brevity, gene, at any time is 1,

then the value of all its output lines at time $T+1$ is simultaneously 1. Thus, the state of the outputs of a gene at $T+1$ depends on its activity at time T alone. For our logical analysis, it is sufficient to allow time to occur in discrete, clocked moments: $T=1, 2, 3 \dots$

A *formal genetic net* is constructed by choosing a value of N , the number of elements comprising the net, and of K , the number of input lines to any gene. Each gene in the net receives exactly K inputs, one from each of K formal genes in N . Inputs arise only from members of N . On the average, each element has K output lines. Nets are randomly constructed in two distinct senses. The K inputs to each gene are chosen randomly; to each gene one of the 2^{2^K} Boolean functions of its K inputs is assigned randomly. After being assembled, these nets are deterministic. We assume that all genes compute one step in one clocked time unit.

Such a genetic net is a finite sequential automaton, a machine with a finite number of states and a function mapping each state into a subsequent state (see Fig. 2). A state of the net is described by a row which lists the present value, 1 or 0, of each of the N elements of the net. Each gene can be independently on or off, thus there are just 2^N distinct states of a net of N binary elements.

If the system is placed in some state at time T , then at $T+1$ each gene scans the present value of each of its K inputs, consults its Boolean function, and assumes the value specified by the function for that input configuration. The net passes from a state to only one subsequent state; therefore, although two states may converge on to a single subsequent state, no state may diverge on to two subsequent states. (The system is state determined.)

There are a finite number of states. As the system passes along a sequence of states from any arbitrarily chosen initial state, it must eventually re-enter a state previously passed. Thereafter, the system cycles continuously through the re-entered set of states, called a *cycle*. The *cycle length* is defined as the number of states on a re-enterant cycle of behavior. A state which re-enters itself, a cycle of length one, is called an *equilibrium state*. Since more than one state may converge on a single state, the state re-entered need not be the arbitrarily chosen initial state. The *transient (or run-in) length* is the number of states between the arbitrarily chosen initial state and the first state encountered on a cycle. A *confluent* is the set of states leading into, or on, a cycle; the *size* of a *confluent* is the number of states comprising it. Each state lies on a single confluent [(see Fig. 2c)].

A formal genetic net must contain at least one behavior cycle; it may contain more. By releasing the net from many different states, each of which runs to only one cycle, the total number of different cycles reached may be counted. The *number of cycles* embodied in a net is the number of different

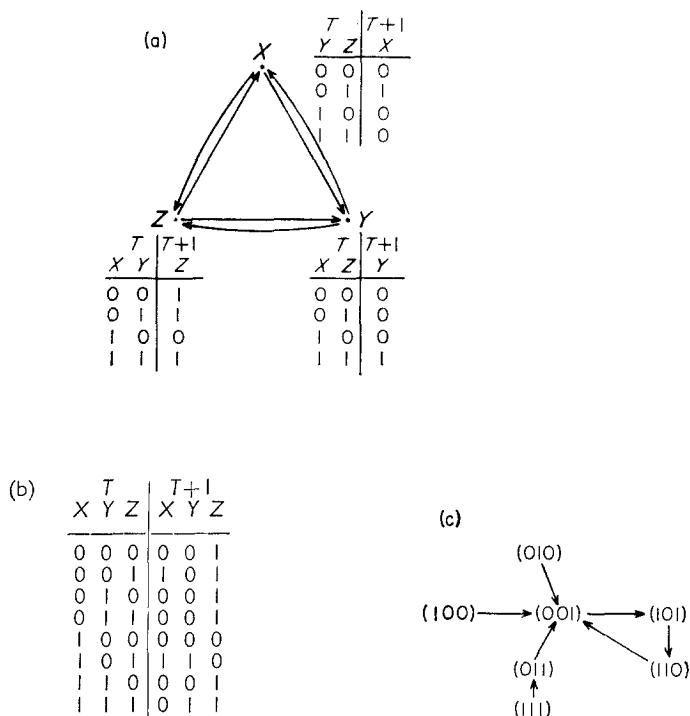


FIG. 2. (a) A net of three binary elements, each of which receives inputs from the other two. The Boolean function assigned to each element is shown beside the element. (b) All possible states of the 3-element net are shown in the left 3×8 matrix below T . The subsequent state of the net at time $T+1$, shown in the matrix on the right, is derived from the inputs and functions shown in (a). (c) A kimatograph showing the sequence of state transitions leading into a state cycle of length 3. All states lie on one confluent. There are three run-ins to the single state cycle.

behavior cycles of which the net is capable. Since no state can diverge on to two subsequent states, no state on one cycle can simultaneously be on a second cycle. Different cycles in one net are behaviorally isolated from one another.

A *distance measure* comparing two states of the net may be defined as the number of genes with different values in the two states. (For example, the state (00000) of a 5 gene net, and the state (00111) differ in the value of three elements.) This distance is used as a measure of dissimilarity between subsequent states on a transient as the system approaches a cycle, between subsequent states along cycles, and between cycles.

As the net passes along a sequence of states on a cycle, one unit of noise may be introduced by arbitrarily changing the value of a single gene for one

time moment. After perturbation, the system may return to the cycle perturbed, or run into a different cycle. In a net of size N there are just N states which differ from any state in the value of just one gene. By perturbing all states on each cycle to all states a distance of one, a matrix may be obtained listing the total number of times the system returned to the cycle perturbed, or ran into any of the other possible cycles. Dividing the value in each cell of this matrix by its row total yields the corresponding matrix of transition probabilities between cycles, under the drive of random, one unit, noise. Such a matrix is a Markov chain. The probability of transition from one cycle to a second need not be identical with the probability of transition from the second to the first. Thus, state noise may induce asymmetric probabilities of transition between the independent behavior cycles of the net.

3. Totally Connected Nets, $K = N$

In random nets in which each element receives an input from all elements, the state subsequent to each state is chosen by sampling at random from an infinite supply of the 2^N distinct states of the net. The characteristics of such a random mapping of a finite set (2^N) of numbers into itself has been solved (Rubin & Sitgreave, 1954). The expected length of the behavior cycle is the square root of the number (2^N) in the set. Therefore, in totally connected nets with 200 elements and 2^{200} states, the expected cycle length is $2^{100} \sim 10^{30}$ states. If the transition from one state to the next required one microsecond, then the time required for a net of 200 elements to traverse its cycle is about 10,000,000 times Hubbel's age of the universe. Totally connected, random nets are biologically impossible.

4. One Connected Nets, $K = 1$

Random nets in which each element receives just one input are no more biologically reasonable than totally connected nets. The structure of a one connected net breaks into separate loops of elements (as in Fig. 2(c) with the direction of all arrows reversed). State cycles arise whose lengths are a maximum of two times the lowest common multiple of the set of structural loop lengths. For random nets as small as 200, the state cycles generally exceed several millions of states in length (Slone, 1967). One connected random nets possess behavior cycles capable of realization by no earthly organism.

5. Two Connected Nets, $K = 2$

The behavior of randomly interconnected, deterministic nets in which each element received just two inputs from other elements is biologically reason-

able. I have studied nets of 15, 50, 64, 100, 191, 400, 1024, 4096 and 8191 elements both by simulation on digital computers and analytically. Nets of 1000 elements possess $2^{1000} \sim 10^{300}$ possible states. The typical net is restricted to cycle among 12 of these states.

The program used constructs a net of size N by random assignment of the two inputs and one of the $2^{2^2} = 16$ Boolean functions to each binary gene. The net is placed in an arbitrary initial state (for example, with each gene switched off) and, at successive time moments, computes its next state. Each of the sequence of states along a run-in is compared with all previous states, and when the present state is identical to a state of the system x moments previously, a cycle whose length is x states has been identified. If undisturbed, the system would cycle through these x states repeatedly.

5.1. CYCLES

Cycle lengths in such nets are exceptionally short. Data was obtained for at least 100 nets at each of several different sizes, and a histogram of the cycle lengths found in each size net was compiled. Figure 3(a) presents a histogram of cycle lengths found in nets of 400 elements which used all 16 Boolean functions of two inputs equiprobably. The distribution of cycle lengths is markedly skewed toward short cycle lengths. Generally, the modal cycle length is less than the median length, which, in turn, is less than the mean cycle length. Here the modal length is 2, the median is 8, and the mean is 98. Equilibrial states (those which successively become themselves) are common.

Among the 16 Boolean functions of two inputs [see Fig. 1(b)], two are tautology and contradiction. An element assigned tautology is switched on regardless of the previous input values. An element assigned contradiction is constantly off. Thus, $2/16 = 1/8$ of the elements in a $K = 2$ random net are foci of constancy. These foci might be thought necessary to produce short behavior cycles. This is untrue. Nets were also studied in which these two functions were disallowed and the remaining 14 Boolean functions assigned equiprobably. The effect is to increase slightly the expected cycle length in nets of any given size and to shift the distribution of cycle lengths in nets of a given size from that found with all 16 Boolean functions. In Fig. 3(b) is the histogram of cycles from nets of 400 elements which used neither tautology nor contradiction. The distribution is still strongly skewed toward short cycle lengths, but the number of cycles of length one (equilibrial states) has decreased. The mode here is 12, the median is 32, and the mean is 209. Deletion of tautology and contradiction has increased the median cycle length in nets of 400 elements from 8 to 32 states. The distribution of cycle lengths is remarkable also in the preponderance of even numbered cycle lengths.

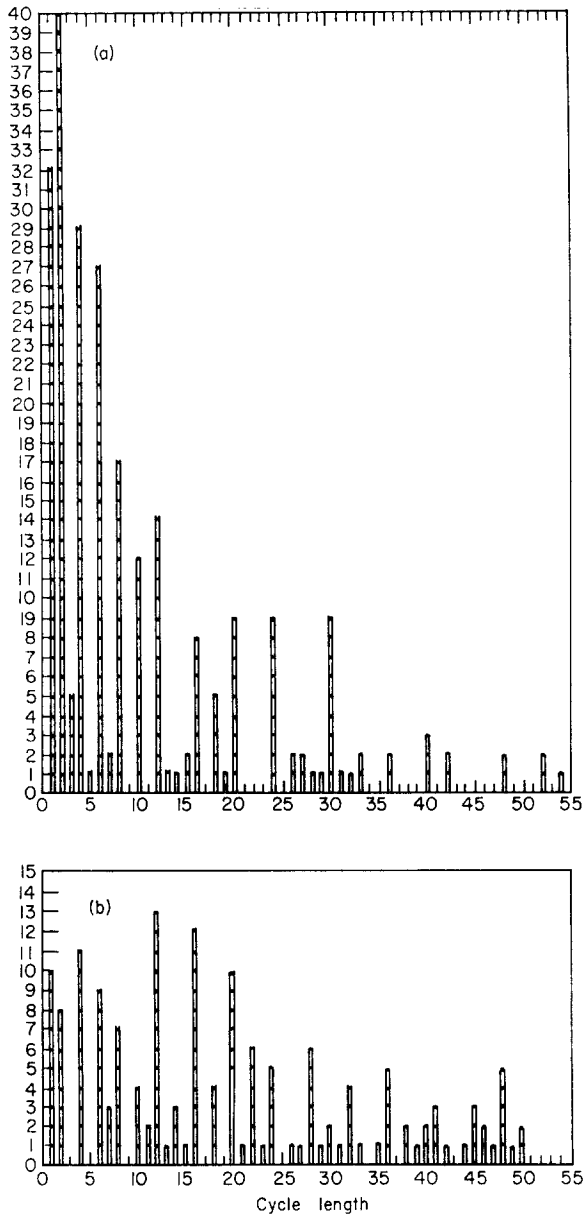


FIG. 3. (a) A histogram of the lengths of state cycles in nets of 400 binary elements which used all 16 Boolean functions of two variables equiprobably. The distribution is skewed toward short cycles. (b) A histogram of the lengths of state cycles in nets of 400 binary elements which used neither tautology nor contradiction, but used the remaining 14 Boolean functions of 2 variables equiprobably. The distribution is skewed toward short cycles.

Because the distribution of cycle lengths is highly skewed, the median cycle length seems the most representational length for nets of any size. In Fig. 4, the log of the median cycle length is plotted against the log of the size net, for nets with all 16 functions, and separately for nets without tautology and contradiction. The values in each condition appear non-linear in the log/log plot. The curves are initially steep, and flatten at larger values of N . In nets with tautology and contradiction allowed, the asymptotic log cycle length against log N is ~ 0.3 . In nets with tautology and contradiction

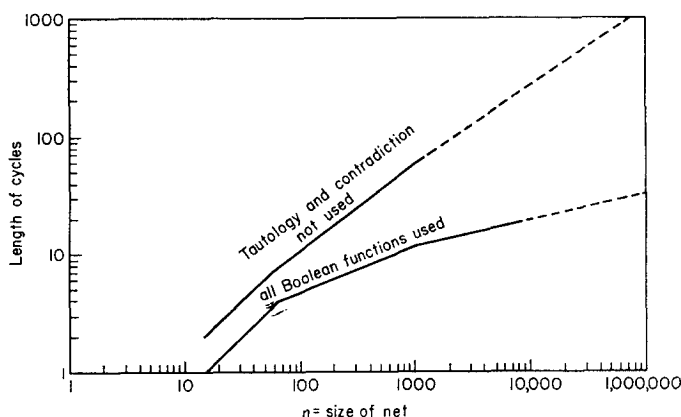


FIG. 4. Log median cycle length as a function of log N , in nets using all 16 Boolean functions of two inputs (all Boolean functions used), and in nets disallowing these two functions (tautology and contradiction not used). The asymptotic slopes are about 0.3 and 0.6.

disallowed, the asymptotic log cycle length $\sim 0.6 \log N$. Disallowing tautologies and contradictions appears to double the asymptotic slope in the log/log plot. In this condition, the expected cycle length is just slightly greater than the square root N (0.5 in the log/log plot). As N increases, the median cycle length initially increases rapidly, then progressively slowly. By projection, nets of 1,000,000 elements, with tautology and contradiction disallowed, possess behavior cycles of about 1000 states in length—an extreme localization of behavior among $2^{1,000,000}$ possible states.

5.2. TRANSIENTS

For nets of a given size, the lengths of run-ins to cycles appears uncorrelated with the length of the cycle to which the transient ran (Fig. 5). The longest transients found were about the same length as the longest cycles found. Like cycle length, the distribution of transient lengths is highly skewed toward short lengths.

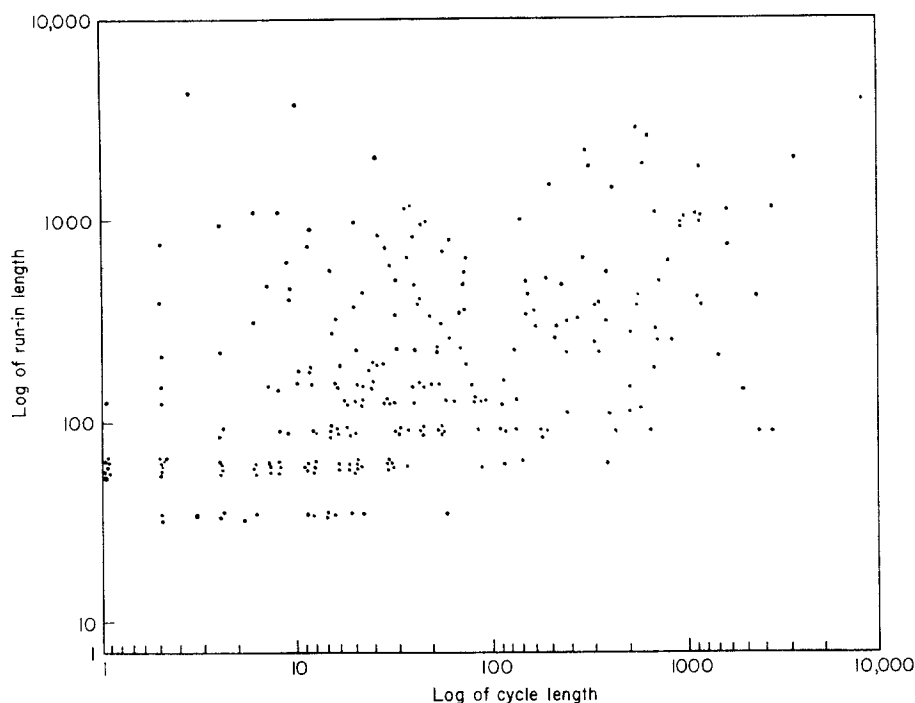


FIG. 5. A scattergram of run-in length and cycle length in nets of 400 binary elements using neither tautology nor contradiction. Run-in length appears uncorrelated with cycle length. A log/log plot was used merely to accommodate the data.

5.3. ACTIVITY

When the system is released from an arbitrary initial state, the number of elements which change value (the activity) per state transition decreases rapidly. In nets of 100 elements, using all 16 Boolean functions, the number of elements which change value at the first state transition is about $0.4N$. This decreases, along a curve nearly fitted by a negative exponential with a half decay of 3–4 state transitions, to a minimum activity of 0 to $0.25N$ per state transition along the cycle. For larger nets, the half decay should require more transitions. Thus, as the system approaches a cycle, states become progressively more similar. One would expect that all states which differ from cycle states in the value of only one element would themselves be located a very few state transitions from that cycle.

The number of genes which change value during a cycle varies between 0 and 35 in nets of 100 elements using all 16 Boolean functions. The consequence is that most genes are constant throughout the cycle, and the cycle states are highly similar.

5.4. NUMBER OF CYCLES

The number of different state cycles—that is, the number of independent and different modes of behavior in these nets—are as surprisingly small as cycles are short.

By computer simulation, nets of 15, 50, 64, 100, 191 and 400 elements were studied. For each net, the system was placed successively in 50 arbitrarily chosen initial states, and the cycle discovered from each initial state was compared with previously discovered state cycles of that net. The median number of cycles per net is low; the distribution of the number of cycles per net around the median is skewed toward few cycles. In Fig. 6 is a histogram of the number of cycles per net, where $N = 400$, and neither tautology nor contradiction was allowed. The median number of cycles per net was 10. Presence or absence of tautology and contradiction does not seem to affect the number of cycles per net.

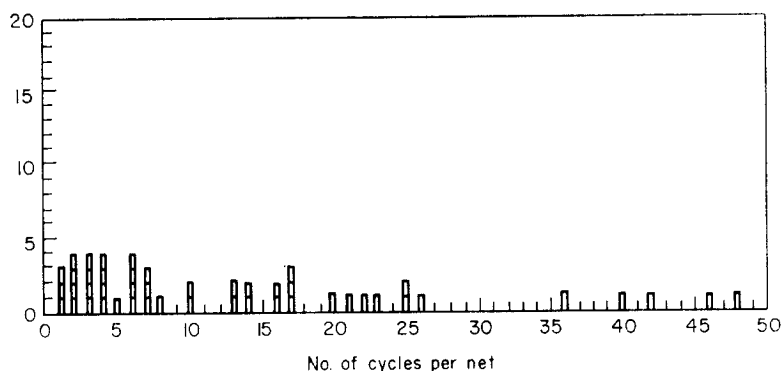


FIG. 6. A histogram of the number of cycles per net in nets of 400 elements using neither tautology nor contradiction, but the remaining Boolean functions of two inputs equiprobably. The median is 10 cycles per net. The distribution is skewed toward few cycles.

The log of the median number of cycles per net is plotted against $\log N$ in Fig. 7. The data appears to fall on a straight line with a slope of 0.5. Log number of cycles $\sim 0.5 \log N$. The expected number of modes of behavior is about $\sqrt{N/2}$. The number of cycles initially rises rapidly, then progressively slowly. By projection, nets of 1000 elements will have about 16 cycles, and nets of 1,000,000 about 500 modes of behavior.

Since only 50 run-ins to each net were made, the data probably underestimates the number of cycles per net. However, 200 run-ins per net rarely revealed more than 10% more cycles than had the first 50 run-ins of the 200; the data in Fig. 7, therefore, seems a good guide for the comparison of the number of cycles per net among nets of different sizes.

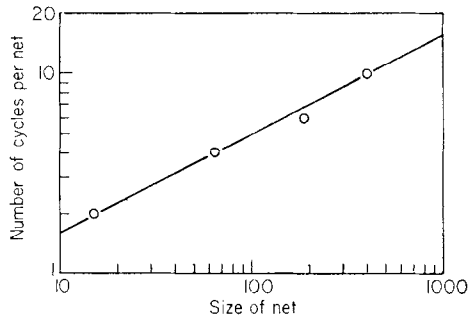


FIG. 7. The median number of cycles per net as N increases appears linear in a log/log plot. The slope is about 0.5. The expected number of cycles is slightly less than square root N .

5.5. DISTANCE BETWEEN CYCLES

The minimum possible difference between states on two distinct cycles is 1—a difference in the value of a single element. This distance occurs frequently but the minimum distance may be as large as $0.3N$. Figure 8 is a scattergram of minimum distances between cycles correlated with the length of the cycles in many nets of 100 elements using all 16 Boolean functions. The median minimum distance between cycles is 5. The average distance between cycles is about 10. When a net embodies many cycles, these frequently form sets within which each cycle is a minimum distance of one from one or two members of the set. Between sets, the distance is larger and may be as great as $0.3N$.

5.6. NOISE PERTURBATIONS

The effect of state noise on the behavior of $K = 2$ random nets has been studied by perturbing the system as it traverses a cycle by arbitrarily reversing the value of a single gene for a single time moment. The perturbed net may either return to the behavior cycle from which it was dislodged, or run in to a different cycle. The program first built a net, then explored it from 50 randomly chosen initial states, and stored the different state cycles discovered. Then all states which differed by the value of one gene from each state of the first cycle discovered were tried, and the cycle to which each of these states ran was stored. From this, a row listing the number of times perturbation by one unit of noise shifted the system from the first behavior cycle to each of the cycles was compiled. The procedure was repeated for all remaining cycles, generating a square matrix listing of the transitions between cycles induced by all possible single units of noise. Division of the number in each cell of the matrix by the row total results in a matrix of transition probabilities under the drive of random (1 unit) noise, which is a Markov chain (see Fig. 10).

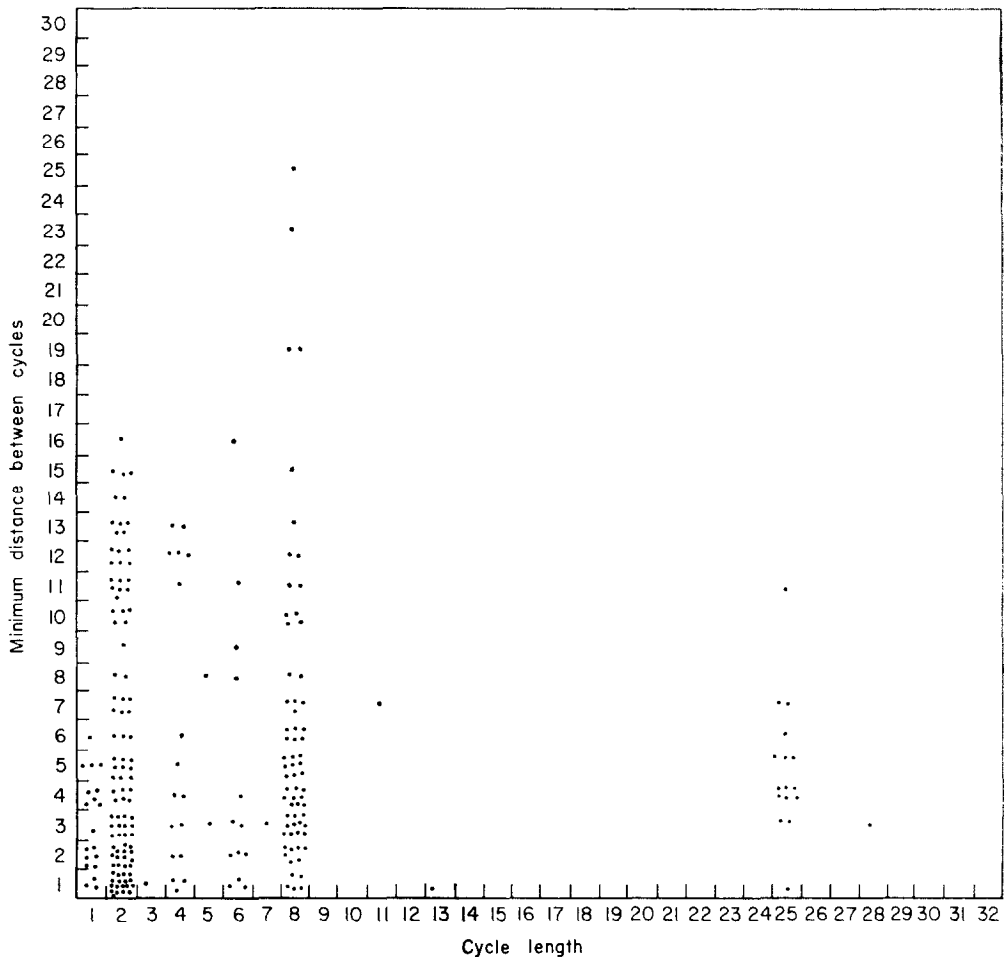


FIG. 8. A scattergram of the minimum distance between cycles and cycle length in nets of 100 elements using all 16 Boolean functions of two variables. Minimum distance between cycles appears uncorrelated with cycle length. The median minimum distance is $0.05N$.

Such chains are characterized by ergodic sets of states, transient states and absorbing states. If each behavior cycle in a binary net is considered a state of a Markov chain, then an ergodic set of cycles is defined to be a set in which each cycle can reach all members of the set by some path through them, but cannot reach a cycle outside the set. A transient cycle lies outside any ergodic set. Once the system reaches an ergodic region, it cannot return to the transient cycle. An absorbing cycle is an ergodic set consisting of a single cycle which always returns to itself after perturbation. Markov chains may,

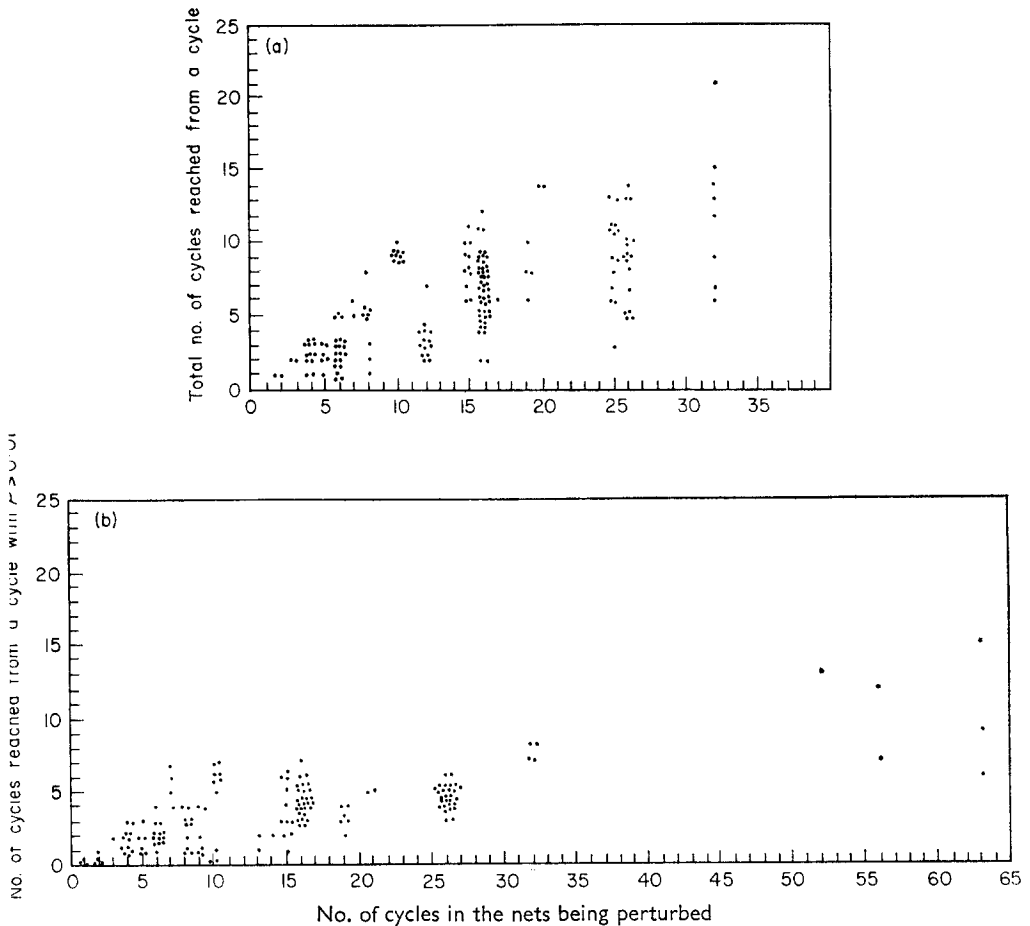


FIG. 9. (a) The *total* number of cycles reached from each cycle after it was perturbed in all possible ways by one unit of noise correlated with the number of cycles in the net being perturbed. The data is from nets using neither tautology nor contradiction, with $N = 191$, and 400. (b) The number of cycles reached from each cycle with a probability greater than 0.01 in the same nets as those of (a). In nets using all 16 Boolean functions, the total number of cycles reached from each cycle is about the same as the data in (b).

of course, have more than one ergodic region; each or all may be accessible from a single or several transient cycles.

Perturbation has been studied in nets ranging from 15 to 2000 elements. Nets larger than 400 elements used all 16 Boolean functions. In those of less than 400, both conditions—with and without tautology and contradiction—were simulated. In general, the net returns to the cycle perturbed with

probabilities between 0.85 and 0.95. Behavior in randomly connected binary nets is highly stable to infrequent noise.

One might have supposed that infrequent noise could induce a shift from each cycle to all others. This proves untrue. Transitions from a cycle are highly restricted; each cycle generally can shift to only one to six other cycles with probabilities of 0.01 to 0.05, and to a few others with probabilities between 0.01 to 0.0001. Most cycles cannot be directly reached from any single cycle [see Figs 9(a) and 9(b)].

Despite the restricted transition possibilities from each cycle in many instances, the entire cycle set forms one ergodic region. Equally frequently, a subset of the cycles forms one ergodic region, and the remaining cycles are transient cycles leading into the ergodic region, but not reachable from it. In the latter case, under infrequent noise, the system may progressively restrict the locale of its activity to the ergodic subset of cycles.

In no case when all possible single units of state noise were explored has more than one ergodic region been found. Restriction of perturbation to the first $0.6N$ of the N genes, however, has on one occasion yielded two ergodic regions. Further restriction of perturbation to $0.05N$ renders multiple ergodic sets probable.

One of the nets studied is presented in Fig. 10. The set of cycles form a single ergodic region, with transients leading into it. One would expect that in systems with several hundred cycles, more than a single ergodic region would be found.

5.7. $K = 3$ NETS

The occurrence of short cycle lengths and few cycles in random nets seems not to depend narrowly on an interconnection of two inputs per gene. I have simulated nets of 15, 20, 25 and 50 elements, each receiving three inputs from other elements, and allowed use of all $2^{2^3} = 256$ Boolean functions of three variables. Cycles were slightly longer, the number of cycles about the same as comparable nets of connectivity two. These characteristic behaviors of random nets seem to require only low connectivity to occur. The rate of their failure as K approaches N will require careful delineation.

6. Discussion

It is surprising that randomly constructed nets, in which each element is directly affected by two others, embody short, stable behavior cycles. The immense restriction of behavior in a $K = 2$ net of 1000 elements, limited to cycles a few hundred states in length, can only be appreciated in contrast to an expected state cycle length of 10^{150} in a totally connected ($K = N$) net of

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	103	2	16			8	1																							
2		289			2		11																							
3	92	20	976		9	2	1					46	5																	
4		49		987			8	32	69				11																	
5	6	24	6		1076												39													
6	46	1	2			324	6										4	1												
7		40					340																							
8		8						329																						
9		49		32			8		987	12									31											
10		3			46					21	966	6			64	8				31										
11		6	27	2						18	9	949	3								85	42	5	1						
12		2	1	46			6					324					3	2												
13		9	18	9									1050																	
14		3			21	46				64				966	8		32								5					
15					40	2	8						2		332															
16						3	6					3	46			316														
17					8	1	7							22	10		322								2		1			
18							11			32																				
19	37		12	2		6			16	9		2						1	939						6	80	39			
20	27		6	18				1	2			86					10			322				2					5	
21			3			3						43																		
22									18	9				35																
23																														
24	3			1		9		6				2					3		9						43	917			106	53
25		12		37	18	2		1	2					9												939	40		6	
26		2		6	1			6																			40	913		
27		6		2			11		1										6	40								33		
28					13			0					45								9				106	9		917	4	53
29																												3	4	303
30							6																	4	3					303

Cycles found
(a)

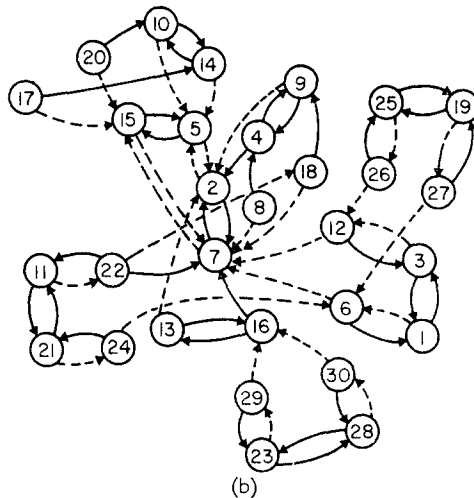


FIG. 10. (a) A matrix listing the 30 cycles of one net and the total number of times one unit of perturbation shifted the net from each cycle to each cycle. The system generally returns to the cycle perturbed. Division of the value in each cell of the matrix by the total of its row yields the matrix of transition probabilities between modes of behavior which constitute a Markov chain. The transition probabilities between cycles may be asymmetric. (b) Transitions between cycles in the net shown in (a). The solid arrows are the most probable transition to a cycle other than the cycle perturbed, the dotted 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* the transitions between cycles are included, the ergodic set of cycles becomes: 1, 2, 3, 5, 6, 12, 13, 15, 16. The remainder are transient cycles leading into this single ergodic set.

the same size. 10^{150} assumes its appropriate proportion when one remembers that 10^{23} estimates the age of the universe in microseconds.

Schrodinger (1944) noted that high molecular specificity, guaranteed by quantum stabilization, is required for the precision of biosynthesis in living things. The behavior of these randomly connected nets discloses an unsuspected, and, I believe, fundamental corollary to that precision. A molecular reaction net of high specificity *is* a net of low connectivity. High specificity appears necessary both for precision of product formation, and to yield a system whose global chemical oscillatory behavior is brief and stable.

The hypothesis that living genetic nets are randomly assembled does not imply that one gene of these nets lacks a specific effect on a second. It asserts that if the "wiring diagram" of the specific repression and derepression connections between genes were known, it would be topologically indistinguishable from a "wiring diagram" generated by random assignment of specific interactions between genes. The hypothesis is consistent with both the random modifications of protein structure induced by mutation, and the lack of steric similarity between the molecule mediating end-product inhibition of an enzyme, and the substrate of that enzyme.

Biologically reasonable behavior in random nets occurs only if each element is directly affected by about the same low number of other elements as are macromolecules in living things. This correspondence lends support to the hypothesis that living metabolic nets are randomly constructed.

7. Cell Cycle Time

Among the most characteristic cyclic phenomena in cells is their replication. Van't Hof & Sparrow (1963) have studied the minimum division cycle time in cells of several species of higher plants. In their Fig. 3 [reproduced as Fig. 11(a)] they show the minimum cell replication time as a function of the DNA content per cell nucleus in six species of plants. The data fall nearly on a straight line. The authors conclude that, in higher organisms, minimum cell replication time is a linear function of the DNA content per nucleus [see Fig. 11(a)].

Projection of this linear function predicts that cells without DNA will require several hours to replicate; bacteria with little DNA per cell require about 30 min to replicate. A curve of replication time from organisms with little DNA per cell to higher organisms must start near the origin, rise rapidly as the amount of DNA per cell increases, then rise more slowly as the DNA per cell continues to increase. Van't Hof & Sparrow (1963) suggest the assumption of a second mechanism to control the time required for cell replication which would provide a steep linear slope from the origin, and

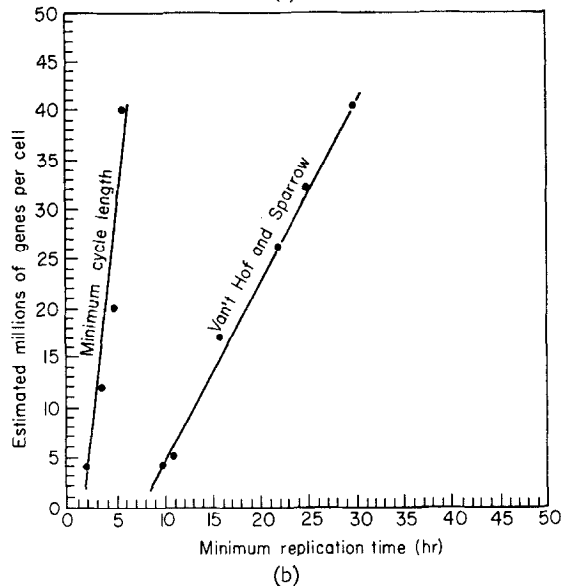
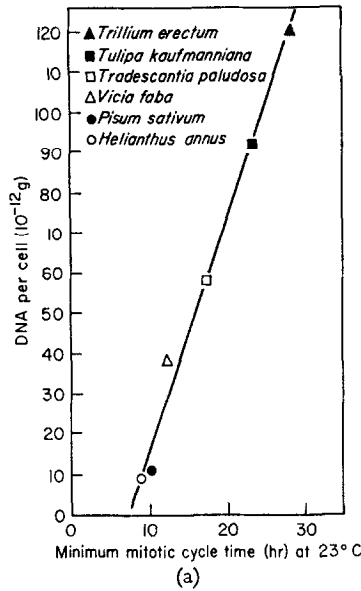


FIG. 11. (a) Van't Hof & Sparrow's (1963) Fig. 3 showing minimum cell replication time as a function of the DNA per nucleus for several plant species. (b) Projected cycle time in nets of 2 to 40 million binary genes using all 16 Boolean functions of two input variables, compared to Van't Hof and Sparrow's plot. In the range where Van't Hof and Sparrow report a linear relation, the binary net model predicts values which are nearly linear. Reduction in the number of elements assigned tautology or contradiction should raise expected cycle lengths and shift the nearly linear slope of the theoretical data to correspond closely with Van't Hof and Sparrow's data.

intersect their observed linear function among higher plants. Choice of control mechanism would depend upon the nuclear content of DNA.

I wish to show that a single principle, the hypothesis that living things are typical randomly interconnected reaction nets, is able to predict cell replication time as a function of the number of genes per cell throughout a wide range of phyla.

Estimates of the time required to switch a gene on or off lie between 5 and 90 seconds (Goodwin, 1963). I will assume that about one minute suffices for a state transition in a real genetic net. Thus, if the model predicts a state cycle length of 100, the biochemical realization of the model should require about 100 minutes to traverse its cycle of oscillatory chemical concentrations.

In Fig. 12 I have plotted the logarithm of cell replication time in minutes against the logarithm of the estimated number of genes in that cell, for several species. The data include bacteria, protozoa, yeast, *Aspergillus*, sea urchin, chicken, mouse, rat, man, rabbit, dog, frog (and minimum cell replication time for) *Vicia faba*, and several other plants (see Table 1). The

TABLE 1
Data for Fig. 12

Organism	DNA per cell	Cell replication time
Bacteria	Watson (1965)	Altman & Dittmer (1962)
Protozoa	Nanney & Rudzinska (1960)	Altman & Dittmer (1962)
Sea urchin	Sparrow & Evans (1961)	Mazia (1961)
Chicken	Vendrely (1955)	Cleaver (1967)
Mouse	Vendrely (1955)	Cleaver (1967)
Rat	Vendrely (1955)	Cleaver (1967)
Man	Vendrely (1955)	Cleaver (1967)
Rabbit	Vendrely (1955)	Cleaver (1967)
Dog	Vendrely (1955)	Cleaver (1967)
Frog	Vendrely (1955)	Cleaver (1967)
<i>Vicia faba</i>	Van't Hof & Sparrow (1963)	Van't Hof & Sparrow (1963)
<i>Pisum sativum</i>	Van't Hof & Sparrow (1963)	Van't Hof & Sparrow (1963)
<i>Tradescantia paludosa</i>	Van't Hof & Sparrow (1963)	Van't Hof & Sparrow (1963)
<i>Tulipa kaufmanniana</i>	Van't Hof & Sparrow (1963)	Van't Hof & Sparrow (1963)
<i>Helianthus annuus</i>	Van't Hof & Sparrow (1963)	Van't Hof & Sparrow (1963)
<i>Trillium erectum</i>	Van't Hof & Sparrow (1963)	Van't Hof & Sparrow (1962)
<i>Aspergillus nidulans</i>	Horowitz & Metzenberg (1965)	Rosenberger & Kessel (1967)
<i>Saccharomyces cerevisiae</i>	Horowitz & Metzenberg (1965)	Williamson (1964)

number of genes per cell was estimated by comparison of its DNA per cell with that of *Escherichia coli*, which Watson (1965) has estimated to have about 2000 genes. Based on these procedures, human cells embody about 2,000,000 genes.

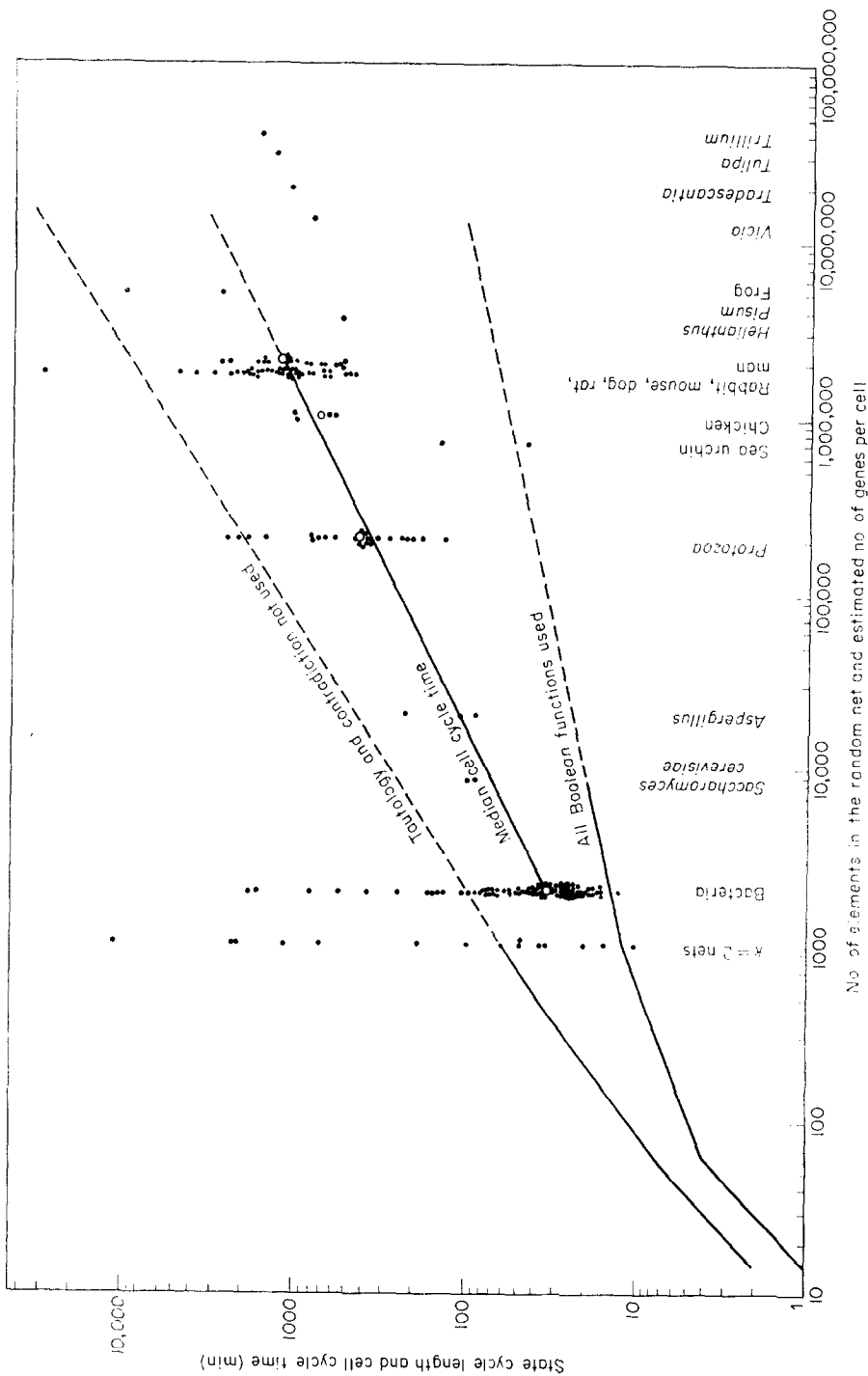


FIG. 12. Logarithm of cell replication time in minutes plotted against the logarithm of the estimated number of genes per cell for various single cell organisms, and various cell types in several metazoan organisms. The data for the plants, *Vicia faba*, *Pisum*, etc. are the minimum replication times described by Van't Hof & Sparrow (1963). The solid line through the biologic data connects the median replication times of bacteria, protozoa, chicken, mouse and dog and rabbit, and man. Data from binary nets of 1024 elements using neither nor contradiction are included for comparison. Median cycle lengths in binary nets with and without tautology and contradiction, as a function of the number of elements in the net, are superimposed on the biologic data. Scale: 2×10^6 genes = 6×10^{-12} g DNA per cell.

The median cellular replication time for bacteria, protozoa†, chicken, mouse, and man are also shown in Fig. 12. It is apparent that these median replication times fall very nearly on a straight line whose slope on a log/log plot is 0.5. The expected replication time in minutes is therefore about the square root of the estimated number of genes. The square root of N increases rapidly initially, then more slowly.

The behavior of randomly interconnected reaction nets predicts this observed relation between DNA content and replication time. The length of state cycles in random nets increases at almost the same rate as a function of the number of elements, as do cell cycle times as the number of genes increases. Based on the assumption that a state transition requires about one minute, the model, without tautology and contradiction, predicts a cycle time of about 50 minutes in a net of 2000 elements, and 16 hours in a net of 1,000,000. The rate of increase of cycle lengths in nets with and without tautology and contradiction are shown superimposed on the biologic data of Fig. 12. Cell replication time falls between the two. In nets using neither tautology nor contradiction, the asymptotic slope of the logarithm of the cycle length is about $0.6 \log N$; using all 16 Boolean functions the asymptotic slope is 0.3. Decreasing the estimate of the time required for a state transition in a real genetic net from one minute to 0.5 minute, brings the theoretical curve for nets without tautology or contradiction into close agreement with the observed slope of log median cell replication times against log number of genes.

In the range of DNA per cell where Van't Hof & Sparrow (1963) describe a linear relation between the DNA content per cell and minimum replication time, the relation between net size and cycle length in nets using all 16 Boolean functions is very nearly linear. The two slopes are of the same order of magnitude [see Fig. 11(b)]. Reduction in the number of elements assigned tautology or contradiction should bring the theoretical slope close to the observed.

The model also appears to predict the distribution of replication times in cells with the same number of genes. Bacteria, with about the same number of genes—2000—concentrate their replication times between 12 and 100 minutes, and scatter them up to 2000 rarely. Random nets of 1000

† Bacteria were assumed to have about the same DNA per cell content and to code for about 2000 genes. In protozoa, the number of genes per cell is difficult to estimate due to the macronucleus. I have treated all protozoa as having about the same number of genes per cell, and estimated this number by dividing the cellular DNA content in *Tetrahymena* by the ratio of macronucleus DNA to micronucleus DNA in *Paramecium*.

I assume the DNA per cell in *Aspergillus nidulans* is about equal to that in *Neurospora crassa*. Rosenberger & Kessel (1967) chose growth media to yield disparate replication times in *Aspergillus* (1.4, 1.8, 3.7, 4.7, 7.0, 9.0 hr). I assume the first three represent relatively normal values.

elements, using neither tautology nor contradiction, concentrate their state cycle lengths between 10 and 100 states, and scatter them up to 2000 to 10,000 rarely. In Fig. 12 are several state cycle lengths in nets of 1000 elements. The distribution is similar to that for bacterial replication times. Both distributions are skewed toward short cycle lengths in a linear plot. A more rigorous test of their similarity lies in the fact that both remain skewed toward short cycles in a logarithmic plot, as shown in Fig. 12.

The single hypothesis that living things behave as typical randomly connected switching nets appears to predict moderately well both the rate of increase in the median replication time as the DNA content of cells increases, and also the distribution about that median of replication time.

Is this correspondence coincidental? Replication of the DNA in higher organisms is known to be initiated at many independent sites. Initiation of replication along any small segment of a chromosome is thought to require the activity of a "replicon", and protein synthesis (Mazia, 1961). If these replicons form elements in the total metabolic net of the cell, depending for their own initiation upon the previous synthesis of other materials, it would not be unduly surprising that the periodicity of their activity, the *S* period, is bound by the periodicity of the entire metabolic net.

Viewing the periodicity of the cell cycle as an expression of state cycles in a randomly connected net may account for the lack of effect upon cell replication time of increasing polyploidy (Van't Hof, 1965). Increasing the number of copies of each gene shifts the expression (of the set of copies) of a gene from a binary variable, when there is only one copy, towards a continuous variable, without altering the connections between or function assigned the genes. The set of copies of a gene would now be capable of a graded output depending upon how many product molecules of its input genes were present. Several arguments (Walter, Parker & Yeas, 1967) suggest that if each element (here element = the set of copies of a gene) in a net realizes a continuous, appropriately nonlinear function (e.g. sigmoid) of its inputs, then the net behaves as though it were comprised of binary devices. In this circumstance, cycle lengths should not be greatly changed by increasing polyploidy.

Unorderly nets in which each component directly affects very few others appear to behave with stability as great as that in living things. States on a cycle are similar to each other; only about 15% of the elements change value during a cycle. The remainder emit a constant output. Even more surprising is the stability shown by random nets to random, one unit noise. In these computer simulations a net was often perturbed from any behavior cycle 4000 times or more. Systems perturbed from a cycle return to that cycle with probabilities of about 90%. While there is little data on the stability of a cell's metabolic behavior to infrequent noise, the behavior of random nets

seems to demonstrate sufficient stability to qualify as a model of cellular stability in the face of biochemical noise.

8. Cellular Differentiation

The principles underlying cellular differentiation remain among the most enigmatic in biology. We are required to explain the spontaneous generation of a multiplicity of cell types from the single zygote, to deduce a natural tendency of a system to become increasingly heterogeneous, then to stop differentiating.

Among the important characteristics of cell differentiation are: initiation of change; stabilization of change after cessation of stimulus; the efficacy of many substances, exogenous and endogenous, as inductive stimuli; a limit of five or six as the number of cell types which may differentiate directly from any cell type; progressive limitation in the number of developmental pathways open to any small region of the embryo; restricted periods during which a cell is competent to respond to an inductive stimulus; the discreteness of cell types, that is, the mutually exclusive constellations of properties by which cells differ; a requirement for a minimal and preferably heterogeneous cell mass to initiate differentiation in many instances, and to maintain it in some; the occurrence of metaplasia between undifferentiated cell types, or from an undifferentiated type to a specialized type, but the lack of metaplasia (the isolation) between specialized cell types; and the cessation of differentiation (Grobstein, 1959).

I believe many aspects of differentiation to be deducible from the typical behavior of randomly built genetic nets.

Cells are thought to differ due to differential expression of, rather than structural loss of, the genes. Differential activity of the genes raises at least two questions which are not always carefully distinguished: the capacity of the genome to behave in more than one mode; and mechanisms which insure the appropriate assignment of these modes to the proper cells. The second presumes the first.

Randomly assembled nets of binary elements behave in a multiplicity of distinct modes. Different state cycles embodied in a net are isolated from each other, for no state may be on two cycles. Thus, a multiplicity of state cycles, each a different temporal sequence of genetic activity, is to be expected in randomly constructed genetic nets. It seems reasonable to identify one cell type with one state cycle. To the extent that this binary model, in which the expression of the "gene" is potentially reversible at each clocked moment, is accurate, it demonstrates the common occurrence of multiple modes of behavior in a genetic system.

If this identification is reasonable, the typical number of cycles in a random

“genetic” net must be of the same order of magnitude as the number of cell types in organisms with the same number of genes.

Estimates of the number of cell types in an organism are hazardous, but the number in man may be placed at about 100; in annelid worms, at 57; in jellyfish, between 20 and 30; in hydra, between 11 and 17; in sponges, about 12–14; in *Neurospora crassa*, 5; in algae, 5; and in bacteria, 2, vegetative and spore (see Table 2). The logarithm of the values are plotted against the

TABLE 2
Data for Fig. 13

Organism	DNA per cell	Number of cell types
Man	Vendrely (1955)	Grobstein (1959)
Sponge	Sparrow & Evans (1961)	<i>Estimated from</i> Borradaile, Potts, Eastham & Saunders (1958)
Jellyfish	Sparrow & Evans (1961)	<i>Estimated from</i> Borradaile, Potts, Eastham & Saunders (1958)
<i>Cenadidia</i>	Mirsky & Osawa (1961)	<i>Estimated from</i> Borradaile, Potts, Eastham & Saunders (1958)
<i>Neurospora crassa</i>	Horowitz & Metzenberg (1965)	Baldwin & Rusch (1965)
<i>Saccharomyces</i>		
<i>cervisiae</i>	Horowitz & Metzenberg (1965)	Baldwin & Rusch (1965)
Algae	Horowitz & Metzenberg (1965)	Baldwin & Rusch (1965)
Hydra†	?	Macklin (1968)
Bacteria	Watson (1965)	Macklin (1968)

† The DNA per cell in hydra presumably lies between sponge and jellyfish.

logarithm of the estimated number of genes per cell in each organism, in Fig. 13. A straight line has been drawn through these values; its slope is 0.5.

The logarithm of the number of independent cycles in a random net is also about 0.5 logarithm of the number of genes. By projection of the data established in Fig. 7, nets with about 16,000 genes (comparable to the sponges) should have about 120 cell types, and man, with an assumed 2,000,000 genes, about 700 cell types. These theoretical predictions are also plotted in Fig. 13.

The rate of increase in the number of cycles in random nets as N increases appears almost identical to the rate of increase in the number of cell types of an organism as the number of genes increases. The theoretical curve is shifted to the left, however, and predicts more cell types than are actually counted.

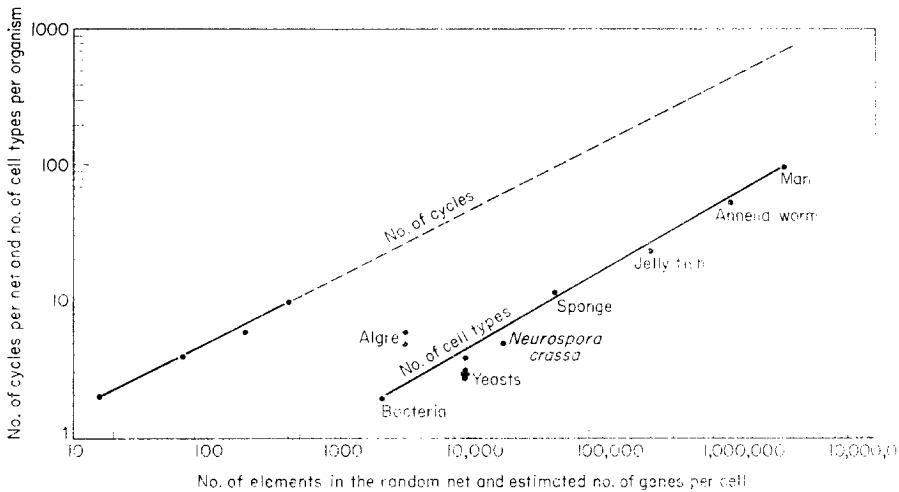


FIG. 13. The logarithm of the number of cell types is plotted against the logarithm of the estimated number of genes per cell, and the logarithm of the median number of state cycles is plotted against logarithm N . The observed and theoretical slopes are about 0.5. Scale: 2×10^6 genes per cell = 6×10^{-12} g DNA per cell.

The predictions remain well within an order of magnitude of the biologic data.

Caution is required for several reasons: large nets have not been simulated; these nets use binary elements, nets of greater verisimilitude must be studied; estimates of the number of cell types in an organism, or the number of genes in that organism are only approximate.

The biologic data need not fall on a straight line in order to remain compatible with this model. Account must be taken of the distribution of the number of cycles found in nets of any given size. The distribution of the number of cycles per net is skewed toward few cycles, as shown in Fig. 6.

With these reservations, I think it fair to say that the correspondence between the predicted number of independent modes of behavior in randomly interconnected nets and the observed number of cell types in organisms is good.

Cells differ from one another in the possession of a constellation of properties which do not intergrade. Similarly, cycles in random nets may be compared for the minimum dissimilarity between their states. In nets of 100 elements, the minimum distance between cycles is commonly $0.05N$ to $0.25N$. Like cell types, behavior cycles are generally separated from one another by a constellation of properties. Since no state may lead to two states, hence be on two cycles, state cycles, like cell types, are mutually exclusive.

If the multiplicity of modes of behavior in a random net helps elucidate the capacity of the genome to behave in more than one way, the appropriate segregation of these modes of behavior to the correct cells requires explanation. Biochemical noise may play a very large role in directing that segregation. A theory which assigns to biochemical noise the task of segregation of different modes of genetic behavior to different cells offers great advantages. Biochemical noise is ubiquitous, unavoidable, and therefore reliable. It remains to show that biochemical noise in a randomly cross coupled genetic net can produce orderly sequential segregation of behavior modes to the appropriate cells.

Perturbation of nets, behaving on cycles, by one unit of noise generally had only a transient effect on the systems behavior. With a probability of about 0.9, the system returned to the cycle perturbed. Of the remaining 0.1 N noisy inputs, these caused the system to shift from any cycle to at most one to six other cycles with probabilities greater than 0.01, and a few more with probabilities below 0.01. It is therefore of considerable interest that, throughout phylogeny, no cell differentiates directly into more than a few other cell types. Restriction in the possible transitions between modes of behavior appears to be characteristic of both random nets and cell types.

The spontaneous generation of a multiplicity of cell types from a single cell type follows explicitly from this model. The occurrence of infrequent noise induces on the cycles of a randomly constructed net the transition probabilities between them which form a Markov chain. Such a chain *must* have at least one ergodic region—a set of cycles each of which can reach all cycles of that set, but no other cycles. It *may* have transient cycles lying outside the single ergodic set, reaching into, but not reachable, from that set. It also *may* have more than a single ergodic set; each ergodic set must be isolated from all other ergodic sets, however, all may be reachable from some single transient either directly, or via other transients.

Let the net embody only a single ergodic set of cycles. If placed on any cycle in the set and perturbed by noise, the system will “spontaneously” pass from cycle to cycle along the allowed transition pathways. An isolated cell would appear to oscillate among its modes of behavior, driven by external noise. If the net is a replicating cell, the clone will explore the permitted transition pathways between cycles and populate the ergodic set according to the asymptotic transition probabilities between the cycles. Cells will spontaneously start to change, pass down restricted pathways of development, populate the complete set of possible modes of behavior, and settle to some stable distribution. Since the net, by hypothesis, embodies only a single ergodic set, each cycle may reach all; differentiation of any cell cannot be stable, a cell of one type should occasionally “spontaneously” change to

become a cell of a different type. Since all types may be reached from any type, deletion of a subset of cell types should create a net movement of cells into the type of cells removed—regeneration should occur.

Stable, irreversible differentiation would require either location in a microenvironment in which the “noise” was sharply biased by the neighboring cell types, or, more fundamentally, a multiplicity of ergodic regions.

Assume a net with more than a single ergodic set. (I have not yet found such a net.) Let there be a transient which reaches, via other transients, into all ergodic sets. Call this transient cycle the zygote. Let it replicate. Then, the zygote is totipotent with respect to all its ergodic regions. We may explain the initiation of differentiation; the zygote is on a Markovian transient cycle in a noisy environment and must eventually leave. We may explain the cessation of differentiation; the system enters an ergodic set and becomes trapped. We have deduced the apparently spontaneous generation of heterogeneity. As the system passes, goaded by noise, from the zygote toward some ergodic set, it must pass branch points to other ergodic sets. Before reaching such a branch point, no noise will move it to that ergodic set; while passing, noise will so move it; after passing, the system will not be competent to respond to noise and reach that ergodic set. Hence we expect the competence of limited duration, the efficacy and reliability of noise as the stimulus for induction, the aid to differentiation provided by contact with heterogeneous tissue as a source of noise; stabilization of change after its induction and the branch point is passed; progressive restriction of developmental pathways as branch points to ergodic sets are passed; limitation in the number of cell types which may arise directly from any cell type; difficulty of metaplastic transformation between specialized cells (in different ergodic sets), and the possibility of metaplastic transformation between undifferentiated (transient) and specialized cell types. Let the net replicate during this perturbation and a particular number will pass down each transition from each cycle, reach each ergodic set, occupy all allowable cycles and distribute themselves according to the asymptotic transition probabilities between cycles and replication rates of each cell type. Grant death to some, and a steady-state population of various cell types arises. Because cells are trapped in separate ergodic regions, overall regeneration is not possible. Within each region, restricted regeneration remains possible. Wounds heal.

Earlier, I alluded to the argument (Walter *et al.*, 1967) that a set of elements whose outputs are sigmoid functions of their inputs behave as a set of binary devices. If true, this suggests that the results obtained for binary nets, rather than being highly simplified approximations, may approach closely to an accurate solution of the behavior of randomly interconnected, biologically appropriate nonlinear, metabolic nets.

Study of the typical behavior of randomly-assembled determinate nets has barely begun. Further research is now needed to extend these results to larger nets; to study the effect of different numbers of inputs per element; to establish firmly the behavior of nets whose elements realize biologically appropriate continuous or probabilistic functions on their inputs; to find the effect of increasing levels of state noise, and more particularly, of "biased" noise due to spatial proximity with other copies of the net behaving on different cycles, and to study the effect of mutation-random alteration in the structure of the net, on its behavior.

While the model has been developed to study cellular control processes, it is formally identical to nerve net models and may find application in other branches of science.

9. Conclusion

A living thing is a richly interconnected net of chemical reactions. One can little doubt that the earliest proto-organisms aggregated their reaction nets at random in the primeval seas; or that mutation continues to modify living metabolic nets in random ways. Evolution, therefore, probably had as its initial substrate the behavior of randomly aggregated reaction nets.

It is a fundamental question whether two billion years of survival pressure have succeeded in selecting from a myriad of unorderedly reaction nets those few improbable, that is non-random and ordered, metabolic nets which alone behave with the stability requisite for life; or whether living things are akin to randomly constructed automata whose characteristic behavior reflects their unorderedly construction no matter how evolution selected the surviving forms.

The data I have presented suggest: that large, randomly interconnected feedback nets of binary "genes" behave with the stability requisite for life; that they undergo short stable cycles in the states of their constituents; that the time required for these behavior cycles parallels and predicts the time required for cell replication in many phyla; that the number of distinguishable modes of behavior of a randomly constructed net predicts with considerable accuracy the number of cell types in an organism which embodies a genetic net of the same size; that, like cells, a random net is capable of differentiating directly from any one mode of behavior to at most a few of its other modes; and that these restricted transition possibilities between modes of behavior allow us to state a theory of differentiation which deduce the origin, sequence, branching, and cessation of differentiation as the expected behavior of randomly assembled reaction nets.

If original proto-organisms built their reaction nets randomly, it behoves the biologist to build an adequate theory of the behavior of these systems;

such a theory should elucidate the problems of biosynthetic organization faced by early living forms. But, if extant biota are also randomly constructed, then an adequate theory of the behavior of randomly assembled reaction nets would constitute an appropriate theory in which to describe the metabolic behavior of nets throughout phylogeny. The correspondence between the behavior of randomly interconnected nets of binary "genes" and the range of biologic data described above, suggest that organisms may indeed form a single population of typical randomly constructed reaction nets. Only if living things do form a single population does a general theory of metabolic behavior seem a reasonable goal for theoretical biology. The consequence of such a theory would be our ability to deduce, not merely describe, metabolic behavior from general propositions about the behavior of any randomly constructed feedback net; and to do so about genetic nets whose exact construction we do not, and may never, know.

Large, randomly assembled nets of binary elements behave with simplicity, stability, and order. It seems unlikely that Nature has made no use of such probable and reliable systems, both to initiate evolution and protect its progeny.

This research was supported under USPHS 5T5 GM 43-05 and USPHS 1-SO-1 FR 5355-04, and partially through the USAEC. A portion of the work was done while a Visiting Scientist at the Research Laboratory of Electronics, M.I.T. from 15th September 1967 to 15th December 1967.

The author wishes to thank Drs Laurel Glass, Sheldon Wolff, Harvey Patt, John Heddle, Creyton Walker, and Warren S. McCulloch for their encouragement and criticism; and Marvin Minsky for making available the use of project MAC computer time while at M.I.T.

REFERENCES

- ALTMAN, P. L. & DITTMER, D. S. (1962). *Growth*.
 APTER, M. J. (1966). "Cybernetics and Development". Oxford: Pergamon Press.
 BALDWIN, H. H. & RUSCH, A. P. (1965). *Ann. Rev. Biochem.* **34**, 565.
 BONNER, J. F. (1965). "The Molecular Biology of Development". London: Oxford University Press.
 BORRADAILE, L. A., POTTS, F. A., EASTHAM, L. E. S. & SAUNDERS, J. T. (1958). "The Invertebrata", 3rd ed. Cambridge University Press.
 BRETSCHER, M. S. (1967). *Nature, Lond.* **217**, 509.
 CLEAVER, J. E. (1967). "Thymidine Metabolism and Cell Kinetics". Amsterdam: North Holland.
 GOODWIN, B. C. (1963). "Temporal Organization in Cells". London: Academic Press.
 GROBSTEIN, C. (1959). In "The Cell", Vol. I., p. 437. London: Academic Press.
 HOROWITZ, N. H. & METZENBERG, R. L. (1965). *Ann. Rev. Biochem.* **34**, 527.
 JACOB, F. & MONOD, J. (1963). 21st Symp. Soc. Study of Development and Growth. London: Academic Press.
 KAUFFMAN, S. A. (1967). *J. Theoret. Biol.* **17**, 483.
 MACKLIN, M. (1968). *Nature, Lond.* **217**, 622.
 MAZIA, D. (1961). In "The Cell", Vol. III, p. 77. New York: Academic Press.

- MIRSKY, A. E. & OSAWA, S. (1961). In "The Cell", Vol. II, p. 677. New York: Academic Press.
- NANNEY, D. L. & RUDZINSKA, M. A. (1960). In "The Cell", Vol. IV, p. 109. New York: Academic Press.
- ROSENBERGER, R. F. & KESSEL, M. (1967). *J. Bact.* **94**, 1464.
- RUBIN, H. & SITGREAVE, R. (1954). "Probability Distributions Related to Random Transformations on a Finite Set". (Tech. Report No. 19A, Appl. Maths. and Stats. Lab, Stanford University).
- SCHRODINGER, E. (1944). "What is Life?" Cambridge University Press.
- SLONE, N. J. H. (1967). "Lengths of Cycle Time in Random Neural Networks". Ithaca: Cornell University Press.
- SPARROW, A. H. & EVANS, H. J. (1961). Brookhaven Symp. Biol. **14**, 76.
- SUGITA, M. (1963). *J. Theoret. Biol.* **4**, 179.
- VAN'T HOF, J. (1965). *Expl. Cell Res.* **37**, 420.
- VAN'T HOF, J. & SPARROW, A. H. (1963). *Proc. natn. Acad. Sci., U.S.A.*, **49**, 897.
- VENDRELY, R. (1955). In "The Nucleic Acids", Vol. II. p. 155. New York: Academic Press.
- WALKER, C. C. & ASHBY, W. R. (1965). *Kybernetics*, **3**, 100.
- WALTER, C., PARKER, R. & YCAS, M. (1967). *J. Theoret. Biol.* **15**, 208.
- WATSON, J. D. (1965). "Molecular Biology of the Gene". New York: W. A. Benjamin, Inc.
- WILLIAMSON, D. H. (1964). In "Synchrony in Cell Division and Growth". (E. Zeuthen, ed.) p. 351. New York: Wiley.