J. theor. Biol. (1971) **31**, 429-451

Differentiation of Malignant to Benign Cells

STUART KAUFFMAN

Department of Theoretical Biology,
University of Chicago, Chicago, Ill. 60637, *U.S.A.*

(*Received* 1 *October* 1970)

Jacob & Monod, and Pitot & Heidelberger have suggested that stable misbehavior of genetic control systems might underlie cancer. Since their papers, considerable insight into the behavior of complex gene control nets has accumulated, and suggests the following. The genome is to be considered a complex dynamic system with a finite number of distinct stable modes of behavior, each of which corresponds to a distinct cell type. Differentiation involves passing among these stable modes of behavior. Regardless of whether higher cells have evolved precise control mechanisms to pass among the stable dynamic modes of behavior, cells are continuously perturbed by biochemical fluctuations which also cause the system to pass among its stable modes of behavior. Such fluctuations are likely to underlie aberrant differentiation. Global features of "aberrant differentiation" in model genetic nets predicts global features of transdetermination experiments. Models and experiments jointly suggest that the normal organism utilizes only a subset of the cell types of which its genome is capable without mutation, thus epigenetic cancer is at least possible. Pathways of differentiation among these normally unutilized cell types exist, and cancer cells might be induced to differentiate along them. Mutation may alter only the pathways of differentiation among cell types, or the cell types themselves. The former case is similar to epigenetic cancer, for cells would follow mutated pathways to reach normally non-utilized cell types of which the non-mutated genome was capable. Such neoplasia might also be induced to differentiate to non-malignant behavior. Even if all cell types are altered by a somatic mutation, not all of them are necessarily malignant–controlled differentiation to such non-harmful cell types might be possible.

Where available, experimental evidence supporting these predictions is examined.

1. Introduction

The majority of research on cancer has concerned itself with understanding the alterations of cells which produce transformations to neoplastic behavior. A relatively small amount of work has been directed towards the question whether cancer cells are capable of differentiating further into non-malignant

modes of behavior, and, if so, to assess the factors which influence such further differentiation.

That relatively little work has been done on attempting to control the occurrence of reversion of neoplastic behavior probably finds its historical explanation in a rather commonly held constellation of beliefs about both cellular differentiation and neoplasia; a set of beliefs which recent experimental and theoretical work suggest is likely to be false.

It appears to be variously supposed that: (1) neoplasia is always due to alteration of the genome by mutation or viral infection; and, once the genome is irreversibly altered, any neoplasia which arises must also be irreversibly neoplastic; (2) even if (as it might be admitted) cancer is possible without alteration of the genome by mutation or viral infection—that is, if epigenetic cancer is possible—nevertheless, any such neoplasia would be an irreversible aberrant differentiation, thus reversion to non-malignant behavior would not be expected. That is, if normal differentiation is irreversible, then so too would any putative aberrant differentiation to yield an epigenetic cancer be irreversible.

There are, however, growing experimental and theoretical grounds to doubt the adequacy of these suppositions. As I shall discuss below, it seems probable that the differentiated state is not always irreversible, therefore it remains possible that epigenetic neoplasia can be induced to differentiate into non-malignant behaviors; furthermore, since cancer cells arising in a genome altered by mutation or viral infection are themselves differentiated cell types of that altered genome, they might be induced to differentiate to benign cell types.

Theoretical grounds for doubting the beliefs noted above grow out of recent mathematical models (Kauffman, 1969) of complex gene control systems which bear on the problem of differentiation and the effects of mutation on the modes of differentiation of higher cells. The models grow out of the Jacob & Monod operon model, idealize the behavior of any single gene as a simple switch, and explore the expected behavior of very large nets of "genes" cross-coupled in arbitrarily complex ways.

Jacob & Monod (1963) themselves suggested that genetic control circuits might be induced to misbehave in stable ways by transient contact with a carcinogen which might, for example, complex a repressor or inducer. Thus, they argued, epigenetic cancer is possible. This thought was also developed by Pitot & Heidelberger (1963). Both pairs of authors devised small circuits based on the operon model which demonstrated ways in which transient contact with some metabolite might alter the system's behavior to a new stable mode of activity. Since that time, considerable work has been done on the expected behavior of very large networks of gene control circuitry which

bears out Jacob & Monod and Pitot & Heidelberger's suggestion that transient perturbation of a single, or a few, gene's activities can cause such a system to change from one to another mode of dynamic behavior, and that epigenetic cancer is possible. These, and other behaviors expected in large genetic control networks will be explored.

2. Model Genetic Nets

The Jacob & Monod operon model requires no introduction. It must be admitted, however, that higher cells may not utilize the same techniques to control transcription, translation, and DNA replication as do bacteria.

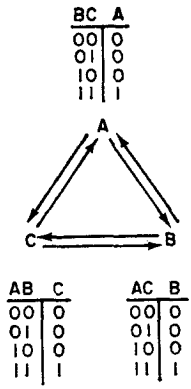
Nevertheless, in higher cells the output of one gene does enhance or inhibit the output of other genes. A mammalian cell may have as many as 1,000,000 distinct genes whose products could influence each others replication and activity. Almost nothing is known concerning the expected behavior of such complex dynamic systems. To attempt to gain insight, it is helpful to idealize the behavior of any single gene as a simple switch capable only of being active or inactive. Surely a gene is not a simple switch, but three claims are made. First, over a short time period, say a minute, a gene presumably is either transcribing, or quiescent. To the extent that transcription (or translation or enzyme activity) is controlled by allosterically modifiable inputs, those inputs can behave as non-linear sigmoid devices (Walter, Parker & Yčas, 1967) roughly treatable as step—that is switch—functions. Finally, while the idealization is an abstraction, I hope it permits the erection of a skeleton of a theory with the right shape.

I model the gene, then, as a binary device, and the genome as a set of N interconnected genes, each receiving inputs from other members of the model genome. In addition to its inputs, each element must be supplied a logical switching function on its inputs which prescribes what each gene shall do at the next time moment, depending upon the current values (1 or 0) of its different inputs. We further suppose that if a gene is on at any time T, then all of its output lines to other genes are also simultaneously, on; and that the genes all compute their next value simultaneously at discrete clocked moments, T_1, t_2, . . . Finally, once the net is built, its structure is fixed.

It is useful to define the following terms: state, state cycle, state cycle length, run-in, confluent and minimum perturbation.

Such a model genome is a finite sequential switching net. A state of the net is a list of the values, 1 or 0, of each of the N components. The net has 2^N possible distinct states. At each clocked moment, each element examines the values of its inputs, consults its logical function, and assumes the value specified for those values of the inputs. Thus, at each moment, the net passes

from a state to a state. Since there is a finite number of states, as the system passes along a sequence of states, it must eventually re-enter a state previously passed. Since the system is deterministic, it must thereafter cycle repeatedly through the re-entrant sequence of states, which is called a state cycle. The cycle length is the number of states on the cycle. A sequence of states not on a cycle, but leading to it is called a run-in. The set of states running into or on a single cycle comprise a confluent. A net must have at least one state cycle but

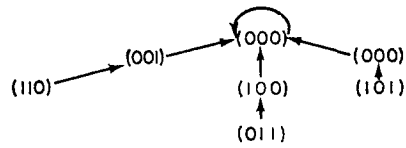


(a)

(Time) T $T+1$

ABC	ABC
000	000
001	000
010	000
011	100
100	000
101	010
110	001
111	111

(b)



(c)

FIG. 1. (a) A net with 3 switching elements, A, B, C, each of which receives inputs from the other two, and switches on only if its two inputs were on during the preceding time moment, T.

(b) on the left, a list of all possible $2^n = 2^3 = 8$ states of the net. For each state, the state to which it goes at the next time moment $(T+1)$ is shown in the right list, and is derived from the net in Fig. 1(a).

(c) Kimatograph derived from Fig. 1(b), exhibiting the sequence of states the system will traverse. State (000) goes to itself on succeeding time moments, and thus is on a state cycle whose length is 1. The maximum conceivable state cycle length is 2^n; state cycles are usually longer than 1. State (111) also goes to itself. The remaining states flow into state (000). The system thus has only 2 state cycles, each comprised (in this particular net) of a single state. The state cycles are the permanent behaviors of the system, the remaining states are transient. All states except (111) flow into or are state (000). These comprise one confluent. State (111) comprises the second confluent.

may have more than one, each a distinct mode of behavior of the entire system.

Since state cycles are permanet repeating behaviors, and run-ins are transient behaviors, the number of distinct state cycles is also the number of distinct permanent ways the system can behave.

3. Orderly Dynamic Behavior

We wish to gain some insight into the sorts of cross-coupled nets of switching components which exhibit biologically reasonable behavior. Although little is known of the cross-coupled interaction between the multitude of a Metazoan's genes, a major clue is available in the high molecular specificity exhibited by proteins. Most known operons have very few direct control inputs—usually two—a repressor and aporepressor, or inducer and repressor (Jacob & Monod, 1963). Similarly, most allosteric enzymes are sensitive to relatively few molecular species of inputs (Monod, Changeux & Jacob, 1963). This suggests that one might classify arbitrarily complex cross-coupled switching systems according to the average number of direct inputs controlling the behavior of components.

Since organisms seem to be comprised of dynamic systems in which each component receives rather few direct control inputs, it is a matter of considerable importance that it is precisely this class of switching systems which typically behave with biologically reasonable order.

A previous publication (Kauffman, 1969), reported the behavior of nets constructed with different average numbers of control inputs per element. Nets in which each element receives direct inputs from every element have exceptionally long state cycles about $2^{n/2}$.† A net with 1000 elements would have state cycles about $2^{500} \simeq 10^{150}$ states in length. Nets in which each element receives just one control input also have very long state cycles. Nets with two control inputs per element have the shortest state cycles, even when the net is constructed entirely at random. State cycles are only about $\sqrt{N}$. Thus, a net with 10,000 elements and $2^{10,000} \simeq 10^{3000}$ states, would typically cycle repeatedly among a miniscule 100 ($= \sqrt{10{,}000}$) states. Almost any net of switching

† Analysis of the factors leading to short state cycles in nets with two inputs per element will be available in 16. That analysis centers around the concept of "forcing structures". Let a binary element A receive an input from B, and B from A, and each receive inputs from other elements of the net. Further let A and B each realize an *or* function on its inputs. Then if A is on, it forces B to be on one moment later, regardless of the values of B's remaining inputs. Similarly, B forces A. The two form a forcing loop which has a stable steady state with both on. Nets with two inputs per element have the shortest state cycles because the size of forcing structures in two input nets is larger than with more than two inputs. Further discussion in this article seems irrelevant.

components—even those constructed randomly—will behave in intensely restricted ways if each element has only two direct control inputs. This result bears upon the fact that cellular chemical systems of high molecular specificity exhibit very restricted dynamic behavior in relation to their possible state space. This supposition is strengthened by the finding (Kauffman, unpublished observations) that nets with three or four direct inputs per element also have very short state cycles, provided a restricted subset of the possible logical functions of three or four inputs are allowed. Restriction to this subset of logical functions is an automatic consequence of the reasonable chemical assumption that inputs to molecular control sites interact only by competing for that site, not by forming aggregates which act as inputs.

To the extent that complex non-linear biochemical systems may be treated as if comprised of switching components, these results suggest that nearly any system built of components showing the high molecular specificity discussed can be expected automatically to exhibit extremely restricted and orderly dynamic behavior, and oscillate in an exceptionally restricted region of its state space.

4. Number of Cell Types

The number of distinct state cycles per net is very low. Nets with two inputs per element ($k = 2$) typically have only $\sqrt{N}$ distinct state cycles per net. Furthermore, computer results on nets with three and four inputs per element suggest that these nets, like $k = 2$ nets, also have only about $\sqrt{N}$ distinct state cycles per net. It appears, then, to be a robust property of nets with very few (but more than one) inputs per element, that they typically behave in only $\sqrt{N}$ ways.

Each distinct mode of behavior of the model genome is a distinct temporal pattern of genetic activity of the genes comprising the system. I will suppose that each distinct mode of behavior comprises a cell type of which the model genome is capable. If the model gives any insight into the behavior of real genetic systems it predicts a relationship between the number of genes in an organism's cells and the number of its cell types. Surprisingly, within the roughness of the data, the number of cell types in an organism is in fact, a square root function of the estimated number of genes per cell (Kauffman, 1969). That there should be even rough correspondence between the theory and data seems worthy of note, in particular since the property of having about $\sqrt{N}$ distinct modes of behavior seems to be a robust characteristic of switching nets with few inputs per component.

These results suggest that the behaviors of nets with two or three inputs per component might serve as useful heuristic models of the gene control system

to provide insight into the processes of differentiation and possible aberration of that process leading to epigenetic neoplasia. The models have revealed two features of complex dynamic systems which appear to be insensitive to the precise construction of the system.

As long as each component has two or three control inputs, then such a system is very likely to behave in very restricted, orderly ways and to have about $\sqrt{N}$ distinct modes of behavior. These systems exhibit several other behaviors which are insensitive to the precise construction of the net, and may explain global characteristics of normal and aberrant metazoan cellular differentiation.

5. Response to Perturbations: Stability

Real cells may well have evolved precise machinery to control their differentiation from one to another cell type. It must be admitted that the evidence for such precise control mechanisms lies mostly in the apparent order of normal ontogeny, rather than upon the elucidation of such precise control mechanisms. In any event, cells cannot avoid biochemical fluctuations. If one conceives of cell types as stable dynamic modes of behavior, then sufficient fluctuations would be expected to cause the system to pass from one to another mode of behavior—that is, to pass among cell types. It is surely likely that random biochemical fluctuations can cause aberrant differentiation, and might therefore underlie carcinogenesis. The behavior of model genetic switching nets when subjected to perturbations may shed light on the behavior to expect from real cells in fluctuating biochemical environments.

In computer similations, the response of large model genetic switching networks to minimal perturbations has been explored (Kauffman, 1969) by reversing the value of a single "gene" while the system was running on a state cycle, then releasing the system. The system will either return to the cycle perturbed, or run into another state cycle. We may perturb a system in all possible minimal ways from each state cycle, and ask how often it returns to the cycle perturbed, how often it goes to each of the other cycles, and to how many cycles the system can go from any given cycle. A matrix summarizing these data as proportions of the total number of possible minimal perturbations is a transition probability matrix exhibiting the chance that random minimal fluctuation will cause the system to shift from the ith to the jth mode of behavior. I will refer to these transitions among modes of behavior induced by perturbation as flows among the modes of behavior. The same global characteristics of these flow paths among "cell types" emerge in almost any switching net with two or three direct control inputs per element. Specifically: a system returns to the cycle perturbed for about 90 to 95% of the perturbations, hence behavior is quite stable to minimal perturbations.

The remaining 5 to 10% of the perturbations of a single element at a time cause the systems to shift to a new mode of behavior. Jacob & Monod and Pitot & Heidelberger's argument that transient perturbation to one component in a very small gene control circuit could shift it to a new stable mode of behavior also applies to almost any very large gene control system. This is not a trivial point for systems can be built whose behaviors are stable to perturbation of any single element.

A single mode of behavior can only flow to a rather small subset (usually two to eight) of the other modes of behavior under the drive of minimal perturbation. When the total number of a net's cycles is greater than about ten, these systems exhibit what might be called *limited local accessibility*, for no mode of behavior can directly reach (flow to) all other modes of behavior under minimal perturbation. Flows between "cell types" may be one way; that is, it is often possible for the *i*th mode to directly reach the *j*th mode, but the *j*th cannot directly reach the *i*th. This is an extreme case of the general characteristic that the probability of flowing from the *i*th to the *j*th state cycle is different from the probability of the reverse flow.

Consider these consequences of dynamic systems which display limited local accessibility.

(1) Non-specific stimuli, affecting many different components of the system, have relatively specific effects on the behavior of the system. Since the system can move from each mode of behavior to only a few others, non-specific stimuli can only perturb it to one of the modes it could reach. Indeed, many different stimuli, acting on *different* components of the system cause the system to flow to the *same* new mode of behavior. That set of stimuli has a perfectly specific effect despite acting on very different points in the system.

This behavior is very reminiscent of the specific inductive effects of very non-specific stimuli during the course of differentiation. It strongly suggests that non-specific stimuli achieve their specific results, not by all acting on some common trigger molecule, but because the system being perturbed in many different ways is constrained to respond in only a few possible ways. A corollary of this is that if cancer can be epigenetic, we should not expect all the stimuli that cause a given type of cancer cell to exert their effects on the same elements within the cellular dynamic system. The specificity is to be found in the restricted modes of response of the cell to whatever stimuli. Thus, even if cancer can be epigenetic, these results suggest that it will not be reasonable to search for some single keystone molecule which all epigenetic carcinogenic stimuli must affect.

(2) Limited local accessibility implies that from a single mode of behavior the system can directly reach only a subset of its other modes of behavior. In turn, this implies that to pass from mode of behavior A to some mode of behavior C, the system must pass through intermediate modes of behavior B. That is, limited local accessibility implies that there must be *pathways* of differentiation among the modes of behavior of a system, with some modes between other modes. In turn, this implies that there will be branch points from which two or more different modes of behavior are reachable, but which are not directly reachable from other modes of behavior. Thus, as a system flows along a pathway it will have limited periods of competence to respond to diverse stimuli by differentiating in a particular way.

It is possible to ask whether, in systems exhibiting limited local accessibility, it is possible to pass from any mode of behavior to *all* other modes of behavior, through other intermediate modes of behavior. That is, one may define the concept of *global accessibility*, and ask whether there is any mode of behavior which is unable to reach *all* the other modes of behavior under the perturbation that is considered here. If there is at least one mode of behavior which cannot reach *all* the modes of behavior, I will say that the system exhibits *limited global accessibility*. If *no* mode of behavior can reach all the modes of behavior through other intermediate modes, I will say that the system exhibits *total limited global accessibility*.

Nets with two or three inputs per element, subjected to minimal perturbations, very often exhibit limited global accessibility. Such systems have a subset of modes of behavior which can reach among themselves, and can be reached by other modes of behavior not in that subset, but cannot themselves reach outside the subset. If we may consider a mode of behavior a reasonable model of a distinct cell type, and flow among the modes of behavior induced by perturbations as a reasonable way of thinking about pathways of cellular differentiation, then the implication is that the system will eventually reach such a subset of modes of behavior (called an ergodic set in Markov theory) and become trapped within it. Or if the system starts within the ergodic set, it will never reach the modes of behavior outside the subset.

Equally often systems of this type exhibit not only limited global accessibility but total limited global accessibility that is, no mode of behavior can reach all other modes of behavior. The probability that a system exhibits either total limited global accessibility or merely limited global accessibility is greatly enhanced if, instead of allowing all the elements to be perturbed one at a time, only a rather small subset of the elements are allowed to be perturbed.

Although nets with two or three inputs per element typically exhibit limited local and global accessibility, these characteristic behaviors are not automatically to be expected in any complex net of switching components. In nets where each element receives inputs from all other elements, there is no difficulty constructing nets in which each state cycle can directly reach several thousand other state cycles by minimal perturbations. These systems exhibit neither limited local or global accessibility among the modes of behavior. Biochemical systems of high molecular specificity, and this few inputs per component, should, however, exhibit these behaviors.

6. Analysis of Experimental Data

These predictions concerning the global features of differentiation induced by biochemical fluctuations find their strongest experimental support in the remarkable work of Hadorn (1967) and Gehring (1968). These investigators have implanted imaginal disks of larval *Drosophilia* into abdomens of adults, maintained the disks for many transfer generations, then retransplanted the disks back to metamorphosing larva. Without the hormones to metamorphose, disks, in adult abdomen, usually remain unchanged except for cell proliferation. The specific questions asked were whether an imaginal disk which normally determined a given structure, say a leg, would again give rise to a leg when reimplanted into a metamorphosing larva. In general, such a disk develops as it normally would, leg disk giving rise to leg structures, wing disk to wings etc. Occasionally, however, a disk is *transdetermined* and gives rise to a structure other than its normal one. Among the most remarkable features of this process of transdetermination is the character of the flow diagram among the structures generated (see Fig. 2). Each structure transdetermines

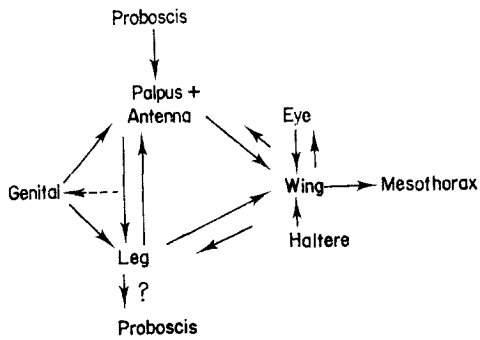


FIG. 2. Comprehensive scheme summarizing the observations of several authors. Short arrows indicate rare transdeterminations. The dotted arrow indicates that transdetermination to genital cells was observed but it is not known from which cell type these are derived (after Gehring, 1968).

directly into only a subset of the other structures. For example, haltere can transdetermine into wing, but not into leg. Transdetermination displays limited local accessibility. The transdetermination of a given structure to a given other structure appears to occur with a fixed probability per transfer generation, but the precise time of a given transdetermination event appears to be otherwise random as would be predicted if the transdetermination were due to random fluctuations.

Some transdeterminations are reversible, but with markedly different probabilities in the two directions; a characteristic exactly parallel to the different probabilities of flow from an *i*th to *j*th mode and the reverse flow, in a model genetic net. Like the model systems, Hadorn's system exhibits limited global accessibility. Some structures, once left, are never returned to, hence the remaining structures cannot reach all the structures. This implies that there is a subset of structures into which the system eventually flows and becomes trapped. In the *Drosophilia* system, thorax acts as such a trap, for other structures can transdetermine directly or indirectly (through intermediate structures) into thorax, but thorax has never been found to transdetermine into other structures. In addition to displaying limited global accessibility the system also exhibits total limited global accessibility. No structure can transdetermine into all other structures; hence, starting with any structure, some structure(s) is unreachable.

The differentiative behaviors of this *Drosophilia* system fit quite well the predictions made from the effects of random minimal perturbations on the behavior of model genetic control systems. Since the sequence of transdeterminations followed in Hadorn's system does not follow in any way the sequence of normal ontogeny, it is clear we are dealing with an aberrant form of differentiation of the sort we might suppose due to random biochemical fluctuations. This supposition is enhanced by Hadorn's discovery (1967) that the probability of transdetermination is enhanced in cells which are proliferating rapidly.

Consider that at cell division, adequate amounts of control substances must pass to daughter cells. If above threshold (or below threshold) amounts are required, if the production of these substances is not perfectly coupled with the cell division rate, and if different control substances are produced at different rates, then gradually the amounts of a few control substances at a time will approach their thresholds and random fluctuations in their distribution to daughter cells at division would perturb the two daughter cells in different ways, a few components at a time. Thus, it seems not unlikely that the flow among differentiated structures observed by Hadorn and co-workers is due to random biochemical fluctuations, and that global features of that flow are to be expected in conditions leading to aberrant differentiation.

7. Epigenetic Cancer

By exhibiting the sort of flow among cell types to be expected during aberrant differentiation due to biochemical fluctuations, both Hadorn's system and the model genetic nets support the claim that epigenetic cancer occurs, supply an interpretation of tumor progression as further aberrant differentiation of neoplastic cells, and suggest that cancer cells might be induced to differentiate to harmless cell types.

The supposition that epigenetic cancer occurs, entails the claim that the normal organism utilizes only a subset of the potential cell types of which its genome is capable without mutation, and that cancer cells occur among the nonutilized cell types. Note that the possibility of epigenetic cancer is not entailed by the claim that transient perturbation can lead to a new stable behavior, as Jacob & Monod, and Pitot & Heidelberger argued, for all the stable behaviors of the system might be the normal cell types of the organism and there might be no additional stable modes of behavior corresponding to abnormal cell types.

The behavior of model genetic control systems with few inputs per element, and also of Hadorn's *Drosophilia* system, however, both support the supposition that the normal organism does utilize only a subset of its potential cell types. Both the model systems and *Drosophilia* transdetermination display total limited global accessibility. In both cases, there is a subset of modes of behavior (structures) which can reach among itself, but cannot reach modes of behavior outside the subset. If started within such a subset, a system would normally never reach outside the subset. Further, no mode of behavior can reach all other modes of behavior. These results suggest that there will be cell types not reachable from any cell type the system starts with, and therefore, that the organism starting from the zygote, utilizes only a subset of its cell types.

Even if *normal* differentiation during ontogeny is controlled by very precise mechanisms, rather than random biochemical fluctuations, it is unlikely that during the normal course of ontogeny, all possible cell types are utilized. Clearly, maintenance of machinery to reach all possible cell types from at least one cell type (the zygote) would be difficult, and would not be expected to occur unless there were an extraordinary selective advantage in doing so. In fact, it is clear that our strongest ground for supposing that machinery exists to reach all possible cell types is the orderly sequence of normal ontogeny itself, but that orderly sequence surely does not require that *all* possible cell types are utilized, only that some occur, and in the right time and place.

These considerations jointly suggest that the unmutated genome is capable

of cell types not normally occuring in the organism, that aberrant differentiation to reach and pass among those unutilized cell types can occur due to transient biochemical perturbations, and that epigenetic cancer could exist among those unutilized cell types. On this account, epigenetic cancer would be an expected, if rather rare feature of cellular differentiation itself. Cancer cells would be expected to display odd patterns of gene activity; a prediction best exemplified by the human oat cell carcinoma of bronchogenic origin which which produces an ACTH-like substance, thus mimicing in part, apparently, the gene activity of pituitary cells.

8. Reversibility and Stability of Neoplastic States

Global features of aberrant differentiation due to biochemical fluctuation displayed by Hadorn's system and model genetic nets, ought also to exhibit themselves in epigenetic cancer. Specifically: differentiation ought to occur among the normally unutilized cell types, and ought to show limited local accessibility. Reversibility should occur (in some cases) but with differing forward and backward probabilities; the differentiation ought to occur with fixed probabilities, but at haphazard times, and ought to be more frequent in rapidly dividing cells which might be expected to be subjected to more telling perturbations. Most of these predictions appear to describe the well-known phenomenon of tumor progression (Faulds, 1969), in which diverse cell types arise apparently at random in the tumor mass and in which both regression and progression appear to occur. On this account, progression would find its common explanation in the further differentiation and selection of cancer cells, rather than always requiring the supposition of a sequence of somatic mutations in the tumor mass. Furthermore when cancer is due to viral or mutational alteration of the genome we would still expect differentiation to occur among the odd cell types of which the mutated genome were capable, hence progression would still seem likely to be due, in most instances, to differentiation, not further mutation.

From this viewpoint, the very occurrence of tumor progression indicates that cancer cells can undergo further differentiation, and might be induced to reach non-malignant cell types. Hadorn's results clearly indicate that the determination of a cell type is not immutable. Data on the reversible loss of capacity to synthesize melanin pigments (Whittaker, 1968), the success in getting nuclei from tadpole gastric epidermis to support normal development when transplanted into enucleate eggs (Gurdon, 1966), and many other studies (Ursprung, 1968), also add growing support to the hypothesis that the differentiated state of many cells or nuclei can be altered. At least if cancer can be epigenetic, the possibility of controlling the

further differentiation of such cells to harmless behavior ought to be extensively explored.

9. Mutations and Perturbations

It seems to be a common, if unspoken, assumption that, when cancer arises in a cell line whose genome is irreversibly altered by mutation or viral infection, then that neoplasia itself must be irreversible. Growing theoretical and experimental data cast doubt on this claim.

Since little is known of the behavior of complex dynamic systems such as cells, little is known about the effects of mutations on the behaviors of such systems. Current studies on model genetic switching nets with two or three inputs per component, however, may have interesting biological implications. We may model the mutation of a gene to produce a useless product by permanently switching off an element of a model net. Switching nets with few inputs per component have short state cycles because many of the components are either permanently active, or permanently inactive. In almost any net, a subset of elements is permanently inactive on all of the diverse state cycles of which the net is capable. Mutation of that net by irreversibly switching off one of these elements, or by removing it from the net, will clearly not alter any of the state cycles of which the system was capable. Hence the "cell types" of which this model genome is capable remain unaltered by such mutations. However, the sequences of states running into the diverse state cycles include states in which the subset of elements that are *inactive* on all state cycles, are active. Irreversible mutation of those elements will alter sequences of states running into each state cycle. The consequence is that when the system is subjected to minimal perturbations, the number of distinct perturbations which cause the net to flow from some mode of behavior to another, may have changed. In short, mutation of elements which are normally inactive on all stable modes of behavior of the system leaves those modes of behavior unaltered, but changes the flow probabilities, under minimal perturbation, among the modes.

The model suggests that there may be a set of genes in real cells which are not active in any stable cell type, but are active during transitions among cell types. Mutations to such genes would be expected to alter the flow paths of differentiation, but to leave unaltered the stable cell types of which the genome was capable. It might be noted that a system of this sort would be highly advantageous in evolution. The occurrence of genes active only during transitions among cell types partially uncouples the control of flow among cell types from the cell types themselves. If workable cell types arose, they might be maintained in evolution, while the situations in which they occurred in organisms might change.

Consider the implications of mutations which alter the flow paths among cell types, but not the cell types themselves. If an organism normally utilizes only a subset of its potential cell types, the effect of such a mutation would be to alter the subset of cell types reachable from the cell in which the mutation occurred. Stimuli which caused the cell to differentiate in one way might now cause it to differentiate in a different way. If cancer cells occur among the normally unutilized cell types of the unmutated genome, the new, aberrant flow paths might lead a mutated cell to differentiate into one of these cancer cell types. Since the normal cell types would have remained unaltered, however, perturbations might be found which could induce the cells of the mutated clone to differentiate into harmless cell types.

There is firm experimental support for the claim that, despite mutation, an altered cell might be able to behave in a non-malignant, even normal way. The same data, support the hypothesis that some mutations leave cell types unaltered, but alter the flow paths among cell types. Gehring (1968) studied transdetermination in aristapedia mutants of *Drosophilia*. "*Aristapedia* transforms the arista and the adjacent part of the third antennal segment into a tarsus. When *Aristapedia* antennal disks are cultured and tested in wild-type hosts, tarsal structures are found. However, after a few transfer generations aristae are also found although this particular mutant (SS^a) never forms an arista *in situ*." Clearly, the aristapedia mutant does not deprive the genome of the capacity to produce aristae, but apparently alters the circumstances of their occurrence. Goldschmidt (1955) investigated the effects of diverse mutations on the phenotype of *Drosophilia melanogaster*. He discovered that if he perturbed a developing wild type embryo at particular times during ontogeny with a temperature shock, that organism developed into an adult with a mutant phenotype despite its normal genotype. Goldschmidt called the phenomenon "phenocopying", for a normal embryo could be made to copy a known mutant phenotype—yet evidenced its normal genotype by yielding normal offspring. Particularly salient to the cancer problem was his discovery that he could administer a temperature shock to a known mutant embryo and cause it to have a *normal* adult phenotype. That the genome was mutant was shown by the production of the expected mutant offspring. Thus, a mutant genome may not entail the necessity of abnormal cellular behavior if the appropriate stimulus can be found. Recent data from Berendes (1968) on the puffing patterns in giant chromosomes in the *Drosophilia* salivary glands show that temperature shocks of the type used by Goldschmidt causes marked alteration in the size of many puffs. If puff size be accepted as a sign of gene activity it seems likely that Goldschmidt achieved his success by perturbing the activity of many genes with his temperature shocks, at particular times during ontogeny.

Three implications of Goldschmidt's work require mention.

(1) Obviously, a mutant genome can support a normal phenotype. The possible consequences for mutational cancer are clear.

(2) The occurrence of phenocopying of normals by mutants and *vice versa*, supports the hypothesis that there are mutations which alter flow paths among cell types, but not the cell types themselves. The hypothesis specifically predicts that perturbations to a normal embryo would cause the system to pass down odd flow paths and mimic a mutant; on the other hand, a mutant with changed flow paths but normal cell types might be perturbed to mimic a normal organism. Were all the stable modes of behavior, cell types, of the genetic control net changed by the mutation, it is not at all clear how transient perturbation with a temperature shock would conceivably allow a mutant organism to copy a normal phenotype.

(3) Goldschmidt's results bear on attempts to prove cancer can be epigenetic. Braun (1969) and Pierce (1967), have both argued forcefully that the occurrence of normal cell types as progeny of a cancer cell prove that the cancer cell itself harbored no mutation but was, instead, an epigenetic cancer. Braun has shown that the crown gall tumor of plants can be induced to differentiate into normal plant structures by serial transplantations of teratomatous cuttings to the growing shoots of normal plants. After four serial transplants, the teratomatous tissue gave rise to normal shoots and leaves, went to seed, and yielded normal plants upon germination. Pierce examined testicular teratocarcinoma derived from a mouse. Clones derived from single carcinoma cells give rise not only to masses of embryonal teratocarcinoma cells, but also to diverse normal somatic tissue. Thus a cancer cell can be the progenitor of normal cells. Despite these striking results, the data fail to establish the claim that the neoplastic tissue was an epigenetic phenomenon and contained no mutation, for Goldschmidt's success in causing known mutant phenotypes to develop into normal adult phenotypes demonstrates that normal cell types can derive from known mutant genomes. It is in fact, unclear that any achievable data could ever establish that a given cancer occurred without mutation.

Although the data of Braun (1969), and Pierce (1967) cannot unimpeachably establish the negative claim that cancer can occur without mutation, the data from Hadorn's (1967) study of transdetermination and from model genetic nets both suggest that an organism utilizes only a subset of its cell types and therefore that epigenetic cancer is possible, and Braun and Pierce's data

surely must count as the best possible direct experimental evidence in favor of the possibility of epigenetic cancer.

Even if a mutation or viral infection were to alter all the stable modes of behavior of the gene control system so that no "normal" cell types remained possible, the altered genome would remain capable of a multiplicity of cell types. There is no *a priori* reason to conclude that all such cell types must behave malignantly, and therefore, it is not unreasonable to suppose that such cancer cells might be induced to differentiate into harmless modes of behavior. The clearest data supporting this supposition comes from the recent work of Rabinowitz and Sachs (1968), who have explored hampster embryo cells transformed by polyoma virus.

The morphology of transformed clones in cultures shows multilayering, attesting the loss of contact inhibitions; cells yield tumors on transplantation to susceptible mice. If grown on a glutaraldehyde treated feeder layer, the morphology of some subclones of transformed cells alters, and these cells no longer give rise to tumors when transplanted into mice. Nevertheless, the viral T antigen is still synthesized by the transformed, but now non-malignant cells. Thus, at least the viral DNA coding for the T antigen remains in the cells, and the cells no longer exhibit malignant behavior.

The effect of culture conditions and tissue interaction on tumor suppression was explored by Rubin (1960), who showed that transformed chick cells infected with Rous sarcoma virus will behave as normal cells if grown in the presence of both high concentrations of fetal serum and normal cells.

These considerations suggest that, whether due to viral infection, mutation or epigenetic misbehavior, cancer cells may be able to differentiate into non-malignant cell types. Further data supporting this thesis derives from the apparent maturation of some neuroblastomas into non-malignant masses of ganglia (Cummins & Rusch, 1966), and from the work of Seilern-Aspang & Kratochwill (1962, 1963) on chemical carcinogenesis in the newt. They found that carcinogenic hydrocarbons, applied to the skin at the base of the tail, produced epithelial carcinoma which metastisized widely and killed the host. Occasionally, after metastasis had occurred, the neoplastic tissue totally reverted to normal cell types. The probability of regression was enhanced if the tail of the newt was removed and its regeneration was occurring.

This reversion represents the further differentiation of neoplastic tissue, as does King's success in transplanting frog kidney adinocarcinoma nuclei into enucleate ova and getting partially normal development (King, 1965).

10. Experimental Perturbation of Neoplastic State to Benign State

If cancer cells differentiate, and occasionally yield non-malignant progeny, the fundamental question must be how, if at all, such beneficial further

differentiation might be induced. While answers are unknown, reasonable clues are present.

Undoubtedly, co-ordinated alteration in the activities of many genes is more likely to lead to a desired differentiative change than random alteration in the activities of one or a few genes. Braun succeeded in causing reversion of malignant behavior of crown gall tumors by grafting cuttings onto the growing shoots of normal plants. Presumably the shoots provided highly co-ordinated stimuli to the activity of many genes. Gurdon (1966) has shown that nuclei of tadpole gastric epithelium can be transplanted to enucleate eggs and yield normal tadpoles. The cytoplasm of the egg presumably supplied highly co-ordinated stimuli to the activity of many genes. Tumor reversion in the newt was enhanced by removal of the tail which by prompting regeneration, might have provided co-ordinated stimuli to many genes. Other experimental approaches to attempt to control the further differentiation of cancer cells are available; for example, co-ordinated alteration in the activity of many genes might be supplied by inductive tissue interactions between neoplastic and heterotypic tissue, and alteration in the temporal order of DNA replication might cause co-ordinated alteration in the activities of many genes (Kauffman, 1967).

Unfortunately, it is not at all obvious how one might attempt to supply such co-ordinated alterations in the activities of numbers of genes to *in vivo* neoplasia.

If co-ordinated alterations in the activities of many genes is difficult to supply, it is possible that essentially random alterations in the activities of many genes, or careful alteration in the activities of a few genes might suffice to induce neoplastic cells to differentiate into harmless cell types.

The extent to which local accessibility is limited depends upon the number of components which can be simultaneously perturbed. Earlier we saw reason to suppose that aberrant differentiation due to random biochemical fluctuations would usually be due to significant fluctuations of only a few components at a time. If so, the sort of aberrant differentiation which might underlie carcinogenesis or progression would exhibit rather marked limited local reachability, in which each cell type would differentiate into only a few others. If the activities of many genes can be perturbed simultaneously, a system would be able to reach from any mode of behavior directly to quite dissimilar modes of behavior. Although cells may not possess machinery to perturb the activity of many genes simultaneously, we are not so limited. While it is difficult to perturb many genes' activity in coordinated ways, it is not difficult to perturb many genes' activity in essentially random ways.

The question is whether or not such wholesale alteration of gene activity, if it could be accomplished, would be expected to induce a neoplastic cell to

differentiate into a less harmful cell type. There are reasons to believe it might.

Accept for the moment the hypothesis that cell types are distinct stable modes of dynamic behavior of a genetic control system, and that biochemical fluctuations must cause flow among these cell types. The behavior of model genetic nets indicates that some of the modes of behavior, and some of the flows among those modes, are far more probable than are other modes and flows. Suppose these results apply to real cells. If, in evolution, cells have achieved machinery in addition to random biochemical fluctuation to control the differentiative flow among cell types, then it would be far easier for that additional machinery to work in concert with the spontaneous behavior of the system under perturbation and insure the same differentiative flows among the same cell types which were probable under the drive of inevitable biochemical fluctuations. Any putative control machinery which attempted to hold cells from differentiating in ways which were likely under random fluctuations would be likely to fail. Thus it is not unreasonable to suppose that, if an organism utilizes only a subset of its cell types and pathways of differentiation among them, that subset is likely to have been molded by evolution to be the most stable to perturbations, and to be the most probable modes and flows of the system under the drive of perturbation.

If so, then unco-ordinated perturbation to the activities of many genes should be more likely to cause the gene control system to differentiate into one of the more probable—hence normal—cell types of which it was capable, than to differentiate into cell types which are usually unutilized, improbable under perturbation induced flow, and abnormal. Goldschmidt's success in causing known mutant embryos to phenocopy normal adults when perturbed with a temperature shock that must have altered the activities of many genes in uncoordinated ways, at least demonstrates that wholesale perturbation of gene activity *can* lead from abnormal to normal behavior. His results do not, of course, establish that such transitions are more probable than are transitions to other abnormal behaviors.

Random perturbation to the activities of many genes is experimentally feasible. Goldschmidt's temperature shocks probably altered the behavior of many genes, so might ionic changes which salt off histones, and so might interference with the temporal order of cellular behavior with periodic electric currents or periodic inhibition of DNA, RNA or protein synthesis at different frequencies.

If it is reasonable to suppose that the subset of cell types and differentiative flows normally occuring in an organism are also the most probable modes of behavior under the drive of small perturbations, then the behavior of model genetic nets indicates that there are pathways of differentiation from the

abnormal unutilized cell types to the normally utilized cell types, which might be followed by altering the activity of only a few genes at a time. The most probable modes of behavior in model nets form the subset(s) of modes which act as traps because within each subset member modes can reach among one another, but cannot reach outside the subset to other modes of behavior. The other modes, however, can reach into that ergodic subset(s). That is, there are pathways to the modes of behavior which inevitably occur under perturbation induced flow. A given mode outside the ergodic subset may not reach directly into the ergodic subset, but might have to pass through intermediate modes of behavior to reach that subset. Thus, even if, under the small amount of perturbation being considered, there is no *direct* path from an unutilized mode to an ergodic subset, a sequence of small perturbations can cause the system to pass through intermediate states to the normally utilized modes of behavior. This suggests that, even if we are unable to supply co-ordinated alteration in the activities of very many genes to induce desired differentiation, nevertheless, an appropriate *sequence* of perturbations to the activity of only a few genes at a time might suffice to induce a neoplastic cell to differentiate through a sequence of intermediate cell types to a non-malignant mode of behavior. In at least one experiment, alteration in the level of a single substrate caused alteration in a tumor's phenotype to a less malignant cell type and growth characteristics; Meins (1969), working with crown gall teratoma lines, found that reducing the level of glutamine in a defined medium caused an unorganized tumor to grow in a more organized, less autonomous fashion mimicing another phenovariant of the tumor.

DNA REPLICATION

Throughout this article we have stressed that no alteration of genome structure or activity is likely to be the unique cause of any type of neoplasm, for a major prediction from model genetic nets is that interference with each of very many diverse components can all lead to the same altered behavior. Nevertheless, we can hardly ignore the fact that most neoplasia is characterized by persistent cell replication. It is surely likely therefore that the neoplastic altered behavior to which many perturbations can give rise is associated with the control of chromosomal replication, a point stressed by numerous authors (e.g. Bullogh, 1965). I consider briefly current information of Metazoan DNA replication in light of our discussion of genetic control nets. As an initial caveat, however, I note that persistent, massive, replication is not necessary for lethal consequences of cancer, for insulinoma may kill while of microscopic size (Pitot, personal communication). Conversely, persistent replication can be associated with normal mature-metazoan cell types, as in the intestinal crypts of mammals.

Mammalian cells contain many replicons, each a segment of DNA which replicates as a unit (Painter, Jermany & Rasmussen, 1966). Painter *et al.* (ibid) estimate that human cells have about 10^4 replicons.

The striking finding of many workers (see Pelling, 1966) is that replicons do not replicate synchronously, but exhibit an asynchronous time sequence. It is at least clear in many cases that different segments of the chromosome complement *complete* replication asynchronously (Keyl & Pelling, 1963). In some cases it is clear that different replicons *initiate* replication asynchronously (Plant, Nash & Fanning, 1966). Particularly interesting was Taylor's finding that the asynchronous order of DNA replication in hampster cells differed between different cell types and seemed fairly stable for each cell type (Taylor, 1960). Thus, if there is a system which controls the sequential order of replicon activity, it must be able to behave in distinct ways in the same organism to yield diverse orders of replication in diverse cell types.

There is a suggestive body of evidence supporting the rather obvious hypothesis that a control net among the replicons may exist and determine the order of replication. Kim, Gelbard & Perez (1968) studied the effect of actinomycin D and cycloheximide on DNA replication. They showed that the initiation of DNA synthesis by replicons at different times in S period was almost immediately inhibited by cycloheximide and inhibited with a delay by actinomycin D. Levels of obvious enzymes such as DNA polymerase and thymidine kinase were not significantly reduced by the treatments. This suggests that ongoing protein synthesis of something other than enzymes is required to initiate replication on new replicons. Cummins & Rusch (1966) report evidence suggesting that in physarum polycephalum, proteins synthesized during early S are needed to initiate replication of late S replicons, and that these proteins are not synthesized if early S DNA synthesis is inhibited. These results are consistent with the hypothesis that when replicons replicate, they transcribe a portion of the replicon, which is then translated to a protein that acts as a signal to initiate other replicons. Rao & Goutcharoff's discovery (Rao & Goutcharoff, 1969) that, in *Physarum*, newly replicated DNA is actually active in transcription also fits such a hypothesis.

If we suppose there to be a network by which replicons switch one another on, and if it is true that different cell types in the same organism replicate in different orders, then we might account for the latter by supposing that several different replicons in the network can serve as initiator loci from which activity spreads throughout the net. Initiation at different loci would yield different orders of replication. Maintenance of a specific order of replication would require persistent initiation at a particular initiator locus, each cell cycle, while quitting the mitotic cycle, would require firm inactivation of all initiator loci in the hypothetical replicon net. Conversely, since normal cells

can persistently replicate, the hypothetical replicon net must be capable of some modes of behavior in which not all initiator loci are firmly repressed.

We have seen that switching nets may have modes of behavior not normally reachable from any given subset of modes. It is not unreasonable to think that perturbations to a replicon net might move it to a normally unused mode of behavior in which one or more initiator loci were not repressed, and which would persistently replicate in an odd temporal sequence. If newly replicated genes are active, as Rao & Goutcharoff (1969) found, then, as I have suggested elsewhere (Kauffman, 1967) alterations in the order of replication might alter the cell's patterns of gene expression. Odd orders of replication might persistently replicate and yield odd cell types: neoplasia.

If it is reasonable to think there is something like a replicon net controlling the order of DNA replication, then it seems reasonable to look for altered order of DNA replication in neoplastic cells, and to attempt to cause back differentiation by trying to alter the order of a malignant cell's replication by, say, periodic pulses of cycloheximid during S period.

A final point. It is sometimes asserted that when a cancer is due to a viral or mutational alteration of the genome, the therapy eventually to be hoped for would be the capacity to correct the genetic defect. Even were this possible, it would still be necessary to control the further differentiation of those genetically repaired cancer cells, for the assertion ignores the capacity of the genome to behave in many distinct ways. While the genome were altered, a cell might have been driven to a mode of behavior from which, after restoration of the genome, it would remain unable to return to normal modes. Means of controlling its further differentiation would have to be found. On the other hand, as we have seen, it may be possible to induce a cancer due to viral infection or mutation to behave in a harmless way without repairing the genetic damage.

In summary cancer cells have been considered a subset of abnormal cell types realizable by epigenetic misbehavior, or viral or mutational alteration of the genome. It seems clear that serious efforts to control the further differentiation of cancer cells are warranted. While techniques to direct such differentiation are wanting, and may prove as elusive as the Grail, rational chemotherapy would seem to require rational epigenetic engineering.

This research was supported in part by the Sloan Foundation.

REFERENCES

BERENDES, H. D. (1968). *Chromosoma* **24**, 418.
BULLOGH, W. S. (1965). *Cancer Res.* **25**, 1683.
BRAUN, A. C. (1969). *The Cancer Problem.* New York: Columbia University Press.
CUMMINS, J. E. & RUSCH, H. P. (1966). *J. Cell Biol.* **31**, 577.

CUSHING, H. & WOLBACH, S. B. (1927). *Am. J. Path.* **3,** 203.
FOULDS, L. (1969). *Neoplastic Development*, Vol. I. New York: Academic Press.
GEHRING, W. (1968). In *Stability of the Differentiated State*, Vol. 1. (H. Ursprung, ed.) p. 136. New York: Springer Verlag.
GOLDSCHMIDT, R. B. (1955). *Theoretical Genetics*. Berkeley: University of California Press.
GURDON, J. B. (1966). *Endeavour* **25,** 95.
HADORN, E. (1967). In *Major Problems in Developmental Biology*. (M. Locke, ed.) p. 85. New York: Academic Press.
JACOB, F. & MONOD, J. (1963). *Symp. Dev. Growth* **21,** 30.
KAUFFMAN, S. A. (1967). *J. theor. Biol.* **17,** 483.
KAUFFMAN, S. A. (1969). *J. theor. Biol.* **22,** 437.
KEYL, H. G. & PELLING, C. (1963). *Chromosoma* **14,** 347.
KIM, J. H., GELBARD, H. S. & PEREZ, H. G. (1968). *Expl Cell Res.* **53,** 478.
KING, T. J. (1965). *Ann. N.Y. Acad. Sci.* **126,** 115.
MEINS, F. (1969). Ph.D. thesis, Rockerfeller University.
MONOD, J., CHANGEUX, J.-P. & JACOB, F. (1963). *J. molec. Biol.* **6,** 306.
PAINTER, R. B., JERMANY, D. A. & RASMUSSEN, R. E. (1966). *J. molec. Biol.* **17,** 47.
PELLING, C. (1966). *Proc. R. Soc. B* **164,** 279.
PIERCE, G. B., JR (1967). *Curr. Topics Devl Biol.* **2,** 223.
PITOT, H. D. & HEIDELBERGER, C. (1963). *Cancer. Res.* **23,** 1964.
PLAUT, W., NASH, D. & FENNING, T. (1966). *J. molec. Biol.* **16,** 85.
RABINOWITZ, Z. & SACHS, L. (1968). *Nature, Lond.* **220,** 1203.
RAO, B. & GOUTCHAROFF, M. (1969). *Expl Cell Res.* **56,** 269.
RUBIN, H. (1960). *Virology* **12,** 14.
SEILERN-ASPANG, F. & KRATOCHWIL, K. (1962). *J. embryol. exp. Morphol.* **10,** 337.
SEILERN-ASPANG, F. & KRATOCHWIL, K. (1963). *Wien. Klin. Wschr.* **75,** 337.
TAYLOR, J. H. (1960). *J. Biophys. biochem. Cytol.* **7,** 455.
URSPRUNG, H. (Ed.) (1968). *The Stability of the Differentiated State*, 2 Vols. New York: Springer Verlag.
WALTER, C., PARKER, R. & YČAS, M. (1967). *J. theor. Biol.* **15,** 208.
WHITTAKER, J. R. (1968). In *The Stability of the Differentiated State*, Vol. 1. (H. Ursprung, ed.) p. 25. New York: Springer Verlag.