

MODELING THE NORMAL AND NEOPLASTIC CELL CYCLE WITH "REALISTIC BOOLEAN GENETIC NETWORKS": THEIR APPLICATION FOR UNDERSTANDING CARCINOGENESIS AND ASSESSING THERAPEUTIC STRATEGIES.

ZOLTAN SZALLASI

Department of Pharmacology, Uniformed Services University of the Health Sciences,
Bethesda, MD, 20814 (szallasi@mx3.usuhs.nih.gov)

SHOUDAN LIANG

SETI Institute, NASA Ames Research Center, MS 239-4 Moffet Field, CA, 94035
(sliang@mail.arc.nasa.gov)

In this paper we show how Boolean genetic networks could be used to address complex problems in cancer biology. First, we describe a general strategy to generate Boolean genetic networks that incorporate all relevant biochemical and physiological parameters and cover all of their regulatory interactions in a deterministic manner. Second, we introduce "realistic Boolean genetic networks" that produce time series measurements very similar to those detected in actual biological systems. Third, we outline a series of essential questions related to cancer biology and cancer therapy that could be addressed by the use of "realistic Boolean genetic network" modeling.

Introduction

Over the past several years overwhelming evidence has been produced that human cancer is caused by multiple genetic changes involving a complex regulatory network of genes and biochemical interactions^{1,2}. As a consequence, effective therapy probably will be achieved only through multiple combinatorial therapeutic targets. Traditional biochemical methods are highly biased towards the identification of a single cause for a particular biological effect. For example, molecular biological screening for oncogenes tend to identify only "full oncogenes", i.e., genes that in a certain phenotypical assay will induce malignant transformation by themselves. Identification of three genes that cause cancer together but which individually do not have transforming ability is a significantly more complicated problem. Amongst other difficulties, if it is assumed that the human genome contains 100,000 genes, one has to start with 100,000³=10¹⁵ transfected colonies, requiring tens of thousands of tissue culture dishes, to assure that all possible gene combinations are represented. This example demonstrates that the identification of biologically important gene or protein combinations will call for novel approaches which go beyond the traditional experimental methods. One of these approaches is building genetic networks that would efficiently model the actual behavior of normal and neoplastic cells, based both on our biological understanding of the cell cycle and the wealth of gene expression data and biochemical measurements that will be soon available due to novel technologies such as DNA chips and DNA microarray^{3,4}.

Genetic networks have been used to model biological systems for several decades⁵. For computational reasons, the only model system that has yielded insights into the overall behavior of large genetic networks (up to 100,000 genes) so far is the one introduced by Kauffman almost thirty years ago⁵. In this, the

expression of genes are treated as Boolean variables, they are either ON or OFF, taking the corresponding value of 1 or 0. The state of each gene is determined by a group of genes and the corresponding Boolean function. For a review on the behavior of Boolean networks, the reader is referred to the comprehensive treatment of the topic by Kauffman⁵ and a concise review by Somogyi and Sniegoski⁶. Despite their obvious limitations, such as the elimination of continuous functions (for example dose response curves), Boolean networks can be efficiently used to build models to study the following questions: (1) What is the overall nature of malignant transformation in terms of genetic networks? (2) What changes maintain cancer? and (3) What will be the level of complexity of efficient therapeutic strategies against cancer?⁷.

This paper deals with two issues: (1) The necessary steps to create Boolean genetic networks that closely resemble the actual biological behavior of cells, which we call "realistic Boolean genetic networks". (2) A series of biologically relevant questions that can be answered using "realistic Boolean networks". These issues have practical implications on understanding carcinogenesis and developing effective therapeutic strategies against cancer.

1. Strategy for Generating Boolean Genetic Networks

Genetic networks can be generated in three well defined steps. First, appropriate variables that incorporate all relevant biochemical and physiological parameters must be identified. Second, the interactions between these variables must be defined and translated into mathematically intelligible connections, (i.e., rule sets or functions between the variables). Third, an efficient computer model must be developed that can produce quantitative results at the level of both the entire system and its constituting parts. Our strategy to build "realistic Boolean genetic networks" is structured along these guidelines.

(1) Genetic network variables: The variables in our modeling are "biological parameters" that are defined as any biologically relevant variable in the cell. It can be the expression of a gene (i.e., the level of a particular mRNA or the amount of a protein), a particular status of a protein such as a biologically relevant phosphorylation or any other posttranslational modification, the localization of a given protein, or the presence or absence of a cofactor, such as steroid hormones. The criterion for a "biological parameter" is that a change in the "biological parameter" alone or in combination with other changes must lead to a change in the status of one or more "biological parameter(s)". Furthermore, all chemically identical entities (for example protein molecules), that share the same function and regulation are included in a single "biological parameter". Traditionally, Boolean genetic networks included only gene expression levels as its variables^{5,6}. This leads to the problem that the rules connecting different genes can change. Let us take as an example the case in which a certain gene will have an effect on another gene only if the protein derived from the first gene is activated by a certain posttranslational modification. The activation of that protein can be time and environment dependent, as is often the case, and the posttranslational activation can happen without any

relevant changes in gene expression patterns. An extended network incorporating all relevant parameters, however, will have rules obeying exclusively the laws of physicochemistry. For example, if a given transcription factor is at the right location and in the proper activation status, then it will bind to its DNA binding site, provided there is no inhibitor present. However, that inhibitor can also be incorporated into the model as a "biological parameter" having an "IF NOT" input on the binding status of our transcription factor. It can be easily seen that the inherent time dependent nature of biochemical rules connecting genes have been transformed into a set of time-dependent variables. For example, if the activation status of a given protein is handled as a variable itself, then the model will not have to deal with the uncertainty of whether the protein is active at a certain state of the cell or not. The obvious advantage of this extension is that it makes the genetic network deterministic. The disadvantage is that it significantly increases the number of variables in the genetic network. Estimating the number of variables in an extended network is the subject of another paper⁴, and requires the estimation of, for example, the number of biologically relevant, distinct states of an individual protein. Here we mention without further proof that the overlapping nature of biochemical rules probably will not allow an increase of variables much larger than one order of magnitude. This will leave us with a network of about one million variables, as opposed to a "pure genetic" network of 100,000 variables. The computational resources required to analyze this extended network will significantly increase. On the other hand several quantitative features of Boolean genetic networks, such as the expected cell cycle length, are proportional to the square root of the total number of genes⁵ (or variables). Therefore these numerical characteristics change only three-fold with one order of magnitude increase in the number of Boolean variables.

(2) Regulatory interactions between variables: The regulatory interactions between "biological parameters" in a Boolean network can be represented as a combination of a directed graph¹ and a set of logical functions. The nodes in the graph are "biological parameters", the directed edges lead from the regulating "biological parameter" to the regulated "biological parameter", and, finally, the logical functions will define the status of a regulated "biological parameter" depending on the status of its regulatory inputs. This graph can be characterized by several parameters: e.g., the average number of regulatory inputs per node, the distribution of regulatory outputs per node, the group of permitted logical, such as analyzing, functions, etc. The overall characteristics of this graph will determine the behavior of the genetic network. Kauffman has found, for example, that unless the number of regulatory inputs and the nature of logical functions are restricted, the genetic network behaves chaotically².

(3) Modeling the behavior of genetic networks: The computational time required to analyze a realistic Boolean genetic network increases with the power function of the number of "biological parameters". The analysis of large networks is

possible using methods such as the bit-packing technique developed by Bhattacharjya and Liang³.

2. Realistic Boolean Genetic Networks

The "state" of a genetic network is a combination of the ON/OFF states of its variables. A deterministic genetic network of a given state, when left alone, will produce a series of consecutive states defining trajectories and attractors (for definitions see 5,6).

We define a "Realistic Boolean Genetic Network" (which we will refer to simply as a "realistic network") as a Boolean network^{5,6} that incorporates all relevant biological variables and has a rule distribution that will produce a time series of consecutive states closely resembling the experimentally measured time-dependent changes of "biological parameters". The experimentally measured time-dependent changes of "biological parameters" can be used as a direct input into the genetic network, or certain overall features can be extracted and then tested whether the actual model produces the same overall features.

As of today, less than 10% of all human genes have been identified⁹, and only a fraction of those were studied in context of the cell cycle. Even this limited amount of experimental data allows estimating certain overall features of the cell cycle. In the first comprehensive study of cycling genes, it was found using serial analysis of gene expression (SAGE)¹⁰, that probably less than 2% of all genes are cycling in yeast¹¹. A review of the literature (see several review articles in Cell, 79, 547-582, 1994 and in Science, 274, 1643-1647, 1996)^{12,13} would suggest a similar estimate for human cells, although no comprehensive measurements have been performed to determine the exact ratio of cycling genes so far. Furthermore, reviewing a large number of cycling genes (see the previously mentioned reviews^{12,13}), it is noteworthy that genes do not show frequent oscillations in their expression levels during the cell cycle. There are several genes that significantly change their expression level during the cell cycle, often increasing from undetectable to maximal level of expression and back, but no gene seems to change its expression more than twice during a complete cell cycle. This raises the question of what characteristics of the genetic network determine the number of cycling genes and the lack of frequent oscillation.

In order to investigate whether features such as the lack of variables frequently changing their status and the actual number of cycling genes are also present in genetic network models, we performed computer simulations on random Boolean networks, which is believed to be the best model currently available for a large number of genes⁵. The focus of Boolean network modeling of the genetic network so far has been on connectivity $k=2$, since random networks with a larger number of inputs per gene, when all possible logical functions are used, tend to produce extremely long state cycle periods that would translate into unrealistically long cell cycles in biology⁵. The rule set we employed was designed to exclude functions which has EXCLUSIVE OR and its variant EQUIVALENT on any two of

⁴ Z. Szallasi (manuscript in preparation)

its inputs. These two functions generate chaotic behaviors producing networks with extremely long cycles⁵.

Our initial results suggested that oscillations during the cell cycle in a $k=2$ network were very frequent. Therefore, we have tested genetic networks with alternative overall features in their directed graph representation (see section 1.2). We have studied models with a higher number, 3, 4, or 5 inputs per gene, using hierarchically canalizing functions. The use of hierarchically canalizing functions produces a smaller cycle length for $k>2$ networks. A k -input Boolean function is hierarchically canalizing if (1) the function itself is canalizing⁶, (2) when the canalizing input of the function takes the non-canalizing value, the remaining function with $(k-1)$ inputs is again canalizing, and (3) when both of its canalizing inputs take their non-canalizing values, then the remaining function with $(k-2)$ inputs is still canalizing, and so on. This canalizing property is recursively true for all the inputs of the function. We have produced a list of all such functions for $k=2, 3, 4, 5$. The number of hierarchically canalizing functions for these k values is 14, 96, 1050, and 15036, respectively. Many of these functions do not depend on one of its inputs. In order to investigate the network behaviors that are distinctly associated with the k -input Boolean functions, we screened out the functions that are constant with respect to one of its inputs. The total number of functions that depend on all of its inputs for $k=2, 3, 4, 5$ is, 8, 64, 736, and 10624, respectively.

The networks discussed so far have a homogenous distribution of regulatory outputs and the mean number of regulatory outputs per "biological parameter" is about 2. In real biological systems this is obviously not the case: certain "biological parameters", such as transcription factors, regulate numerous other "biological parameters", while others regulate only a few or none. We have generated a $k=2$ network with canalizing functions in which the mean number of regulatory outputs for half of the "biological parameters" is between 0 and 2, and for the other half is between 10 and 12.

Our results show that oscillations during the cycle are very frequent in all of the networks discussed so far. The majority of all cycling genes change their expression states four or more times when k takes the values of 2, 3, 4 or 5 or when $k=2$ and there is a non-homogenous regulatory output distribution. The exact percentage of cycling genes undergoing this behavior depends on the length of the cell cycle attractor and ranges between 50 and 95% of all cycling genes. Figure 1 shows a detailed distribution histogram for the $k=2$ network with 200 genes, demonstrating that the typical number of expression level changes among the cycling genes is about the half of the cycle length. This is also true for larger networks with $N=1000$ and $N=5000$ (data not shown). Our results for $N=200$ networks with higher k values ($k=3,4,5$) showed that although there is a gradual reduction in the frequency of gene oscillation, this reduction is relatively small and

⁶ A function is canalizing if at least one of its inputs (called canalizing input) has the property that, when it takes a specific value (called canalizing value), the output of the function is independent of all remaining inputs.

the rate of Boolean variables changing their status four or more times per state cycle never drops under 50% of all cycling variables. We also found that the percentage of the cycling genes is relatively large, being 13% for $k=2$ and 20% for $k=3, 4, 5$ even for the shortest period between 4 and 8 transients. The percentage of cycling variables increases with the cycle length. Increasing the value of k from 2 to 5, or changing the distribution of regulatory outputs, did not ensure either a less than 2% rate of cycling genes or a reduction in the number of genes changing their expression status more than twice during the state cycle. These results suggest that in real, biological genetic networks, there must be a special subset of logical rules that will produce the above described overall features.

Currently we are developing algorithms that can generate directed "genetic network" graphs with other desirable features, such as predetermined regulatory output distributions, selected average circuit sizes, and asymmetric logical function distributions.

Large scale gene expression measurements will soon allow the determination of other overall features of the cell cycle that can be used to "update" the overall rule distribution of the "realistic network". These features include (1) the number of differentially expressed genes between alternative cell cycle attractors (i.e., various differentiated cells) and (2) the overall time distribution of cycling genes and parameters.

3. Using "realistic networks" for modeling carcinogenesis.

Malignant transformation was modeled as a "phase transition" of the state cycle attractor of Boolean genetic networks by Kauffman¹⁴. The extent to which Boolean genetic networks represent the actual cell cycle has not been addressed experimentally yet. In this section, we outline the expected changes in large scale expression patterns during malignant conversion depending on the nature of transition in the underlying genetic network. Results from the literature supporting different classes of malignant transformation are also listed.

Let us consider carcinogenesis as a transition from the normal to the neoplastic cell cycle. In terms of genetic networks, there are three possible mechanisms of this transition, each of which may represent real tumorigenesis.

(1) The cell cycle attractor of normal and neoplastic cells may essentially be the same. Transformation by an oncogene induces permanent changes in the state of downstream genes and rules without driving the system from one state cycle attractor to another one. In this case the downstream changes depend entirely on the inducing oncogene. If the oncogene is turned off, the downstream changes immediately go back to their preneoplastic state, and one would expect the cell to display a normal phenotype again. In fact, this was demonstrated recently by the induction of reversible tumorigenesis in mice by conditional expression of the HER2/*c-erbB2* receptor tyrosine kinase¹⁵. Using a tetracycline repressible promoter, Beckers and coworkers have overexpressed this oncogene and transformed NIH-3T3 fibroblasts by withdrawing the suppressor tetracycline. Upon readition of tetracycline, the expression of the oncogene was suppressed, and the cells reverted to

the non-neoplastic phenotype. Whether neoplastic transformation by a single oncogene is the basis of some forms of human cancer remains to be established. It remains to be seen whether conditional overexpression of other oncogenes leads to reversible transformation as well. The therapeutic implication from this model is that these types of tumors, if they exist, could be treated by targeting the "master switch" oncogene. Parallel measurements of a large number of "biological parameters" would reveal the following features in these tumors: (a) most or all differentially expressed genes or other biological parameters in the tumors are permanently ON during the neoplastic cell cycle; (b) the time dependent expression pattern of all cycling parameters, and the actual length of the cell cycle is the same or similar in the normal and tumorous cells; and (c) knocking out a differentially expressed tumor gene will have a permanent effect on the downstream regulated elements and the pattern of permanent changes could be used to work out the hierarchical order of oncogene activated genes, eventually leading to the identification of the transforming gene. In a variation of this model, a combination of two or more permanently expressed oncogenes would induce transformation without driving the cell into an alternative state cycle attractor. In this case the identification of the oncogenes might be somewhat more difficult but the overall features of the network should remain the same.

(2) The second mechanism by which a single oncogene might cause transformation is by forcing the system to leave one state cycle attractor and flow into another one. Once the genetic network enters the attractor basin of a different state cycle attractor, it might undergo a profound reorganization and, without intervention, might unavoidably fall into the new state cycle. The length of this new state cycle may be significantly shorter than the previous one (i.e. faster cell cycle), it may be more resistant to outside perturbations, and less prone to enter resting, G0 - type periods. The expected experimental features of these tumors are: (a) there will be several new cycling parameters appearing in the neoplastic cell cycle; (b) the time dependent expression pattern of all cycling parameters will be different between normal and tumorous cells, as it is reflected by cluster analysis of coregulated genes; (c) since the neoplastic cell is completing the cell cycle by different intermediate states, the number of states and the length of the cell cycle is likely to be different from that of the normal cells; and (d) very often knocking out a differentially expressed tumor gene or inhibiting another biological parameter will not have any effect on the expression pattern of other genes. As for therapeutic implications we are currently addressing the issue by using "realistic models" to determine whether removing the oncogene can drive the system back from the neoplastic state cycle attractor to the normal one, or this will force the system into a third, different attractor. Our results will indicate if targeting the transforming oncogene could provide a cure for these types of tumors as well. Experimental results suggest that these tumors might not be reversible by removing the oncogene. For example, the loss of the *ras* oncogene does not lead to the reversal of neoplastic phenotype in human tumor cells¹⁶.

¹⁶ Z. Szallasi and S. Liang (manuscript in preparation)

(3) The third mechanism involves a series of perturbations. This model does not require the presence of a "master switch", or full oncogene. Instead, it assumes that a series of perturbations in the biological parameters, the genetic basis of which can be termed as "partial oncogenes", will drive the system into the basin of attraction of a new state cycle of neoplastic features. The individual perturbations may not cause any phenotypical change by themselves. The quantitative features of the cell cycle of this type of neoplasia will be very similar to the one described in point 2.

Here we would like to emphasize that different cancer models produce different quantitative features that could be readily tested by today's large scale quantifying technologies such as DNA chips. Therefore, the predicted features could soon be tested. In fact, for yeast, the DNA chip covering its complete genome is already commercially available from Affymetrix¹⁷. Using this methodology one can easily quantitate the expression level of all genes along the yeast cell cycle with the necessary time resolution.

The three possible mechanisms of carcinogenesis outlined above represent fundamentally different etiology and probably call for fundamentally different therapeutic approaches as well. In the second and third mechanisms, the complexity of rearrangements of gene regulatory pathways during carcinogenesis can not be assessed by traditional methods focusing on one or a few regulatory pathways, oncogenes, etc. In these, the whole regulatory network becomes rearranged. Modeling with "realistic models" can give us the first approximate answers for a series of questions with practical implications for cancers involving cell cycle attractor transitions.

The first group of modeling questions are related to the nature of oncogenes and to the identification of "partial oncogenes". Does the stability of cell cycle attractors increase with increasing k values? How many changes in the states of biological parameters are needed to induce a cell cycle attractor transition? How can we identify biological parameters that individually do not induce cell cycle attractor transition but together drive the cell into the basin of attraction of another cell cycle attractor?

Results obtained by modeling for the second group of questions will have an impact on the molecular diagnosis of cancer. What is the number of differentially expressed genes or biological parameters between two cell cycle attractors, where the second attractor is the result of a transition induced by changes in a certain number biological parameters in the first attractor? Furthermore, how many changes can a cell cycle attractor tolerate amongst its cycling and non-cycling parameters without moving to another attractor? The numerical results answering these questions will provide an estimate for the expected number of differentially expressed genes in neoplastic versus normal tissue.

The third group of questions deals with therapeutic strategies. If we drive a cell cycle attractor to an alternative attractor by a change in one or more biological parameters, can we drive the system back to the original attractor by reversing the same changes that induced the first attractor transition? In medical terms: What is our chance to reverse cancer by knocking out the oncogene(s). We can also estimate,

by modeling, the number of changes in biological parameters that are necessary to drive the neoplastic cell cycle into an alternative, preferably normal, cell cycle attractor. Is there any correlation between the average number of differentially expressed genes/ biological parameters in cancer and the number of required changes to drive the cell cycle attractor to an alternative attractor? In other words, these modeling efforts will provide an estimate for the expected complexity of effective therapeutic strategies.

In the previous section, the term "gene" and "biological parameter" were often used side by side, although the group of genes is a subset of all biological parameters. A model incorporating all biological parameters is obviously preferable but since the current technology allows massive parallel measurements of only gene expression, we discussed how the above issues can translate into changes in gene expression patterns.

This leads us directly to the next important issue. Gene expression is only one manifestation of "biological parameters". Depending on the actual rule distribution of the living cells, rearrangements may occur at any level of the "biological parameters". Consequently: We might find most cancer related changes (a) mainly in gene expression, (b) mainly at another level of "biological parameters" or (c) at a combination of the two. A rigorous comparison of the regulation of gene expression, protein expression, protein activation, etc., might reveal different populations of biological parameters, that may be involved in carcinogenesis to a different extent.

The practical implications of this problem are evident. One of the fastest developing fields in molecular biology, for obvious reasons, is the massive parallel quantitation of expression levels and detection of mutations in mRNA samples. Large scale quantitation of protein levels, and especially the simultaneous measurement of a large number of biochemical reactions, are far behind in development. For example, within a reasonably short time it will be possible to quantify the entire transcriptome (all transcribed genes!) with sufficiently high time-resolution of the human cell cycle. However, information about changes in protein levels, posttranslational modification states, etc., will be more limited. Therefore, our chances to identify important cancer associated changes will be high if those are mainly manifested at gene expression or gene mutation levels. If the rule distributions regulating different groups of biological parameters, i.e., mRNA levels, protein levels, etc., are not identical, then "realistic modeling" may give useful estimates about the distribution of cell cycle attractor transition-associated changes. The actual value of these estimates might help cancer researchers and the biotechnology industry to make strategic decisions about the efforts invested into different approaches of large scale measurements.

We believe that developing therapeutic strategies for complex diseases will require computational modeling approaches. In this paper, we have outlined a general strategy how "realistic Boolean genetic networks" can be constructed. Choosing an appropriate set of biological parameters, these models will provide time dependent expression patterns that could be compared to experimental data. This will help with the identification of restricted rule sets that will save significant

computational power, bringing the model into the realm of realization. Once reliable "realistic models" are built, they can be used to assess the complexity of regulatory rearrangements in cancer, and they can help to test multiple combinatorial therapeutic target strategies.

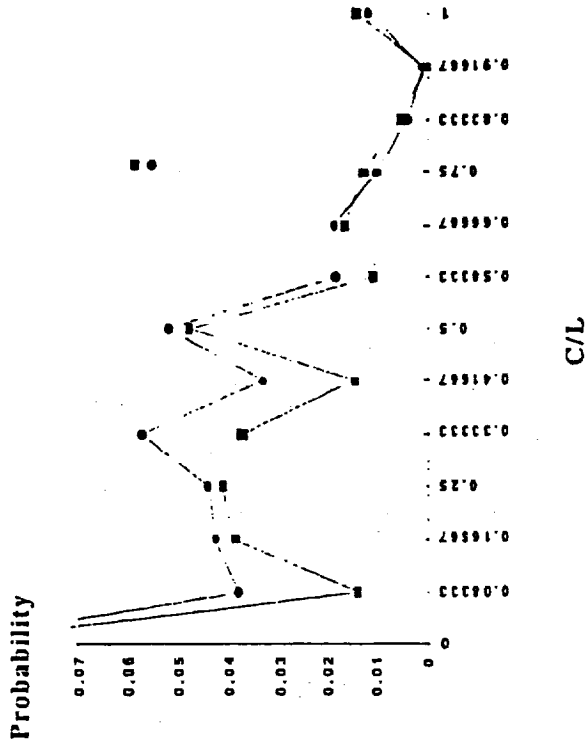


Figure 1. Distribution of probability that a variable changes its value 2, 4, ..., L times during a complete state cycle, where L is the length of the complete state cycle. On the horizontal axis we plotted the number of times (C) a variable changes its value during a complete state cycle normalized for the length of the complete period (L). On the vertical axis we plotted the probability of the given number of direction changes. The figure shows that the number of direction changes peaks around the half of the cycle length. The random network has 200 Boolean variables with two-input per variable ($k=2$) constructed from 8 $k=2$ analyzing rules that depend on both of its inputs. The histogram is obtained by averaging over 3000 networks. For each of the networks we find cycles starting from 500 randomly selected initial state for the Boolean variables. The two curves are obtained from averaging over cycle period from 8 to 16 (filled squares) and from 16 to 32 (filled circles). For these two cases, the percentage of Boolean variables that do not change in the cycle are 74% and 67%, respectively.

Acknowledgment

The opinions and assertions contained herein are the private opinions of the authors and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences or the U.S. Department of Defense. S. Liang acknowledges the support of NASA Cooperative Agreement NCC 2-974. We thank Dr. Gregory T. Kius for critical reading of the manuscript.

References

1. T. Hunter, "Oncoprotein networks" Cell 88, 333 (1997)
2. B. Vogelstein and K. W. Kinzler, "The multistep nature of cancer" Trends Genet. 9, 138 (1993).
3. M. Chcc, et al., " Accessing genetic information with high-density DNA arrays" Science, 274, 610 (1996)
4. M. Schena et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270, 467 (1995)
5. S. A. Kaufman, The origins of order. Self-organization and selection in evolution (Oxford University Press, New York, 1993)
6. R. Somogyi and C.A. Sniegowski, "Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation" Complexity 1:45 (1996)
7. Z. Szallasi, "Genetic networks, carcinogenesis and cancer therapy" Mol. Carcinog. (submitted) (1997)
8. Bhattacharjya, A. And Liang, S. Power-law distributions in some random Boolean networks. Phys. Rev. Lett. 77:1644 (1996)
9. G.D. Schuler et al., "A gene map of the human genome" Science 274, 540 (1996)
10. V.E. Veleulescu et al., "Serial analysis of gene expression" Science 270, 484 (1995)
11. V.E. Veleulescu et al., "Characterization of the yeast transcriptome" Cell 88, 243 (1997)
12. Reviews by several authors in Cell, 79, 547-582, (1994)
13. Reviews by several authors in Science, 274, 1643-1647 (1996)
14. Kauffman, S.A., "Differentiation of malignant to benign cells" J. Theor. Biol., 31, 429 (1971)
15. S. Baasner et al., "Reversible tumorigenesis in mice by conditional expression of the HER2/c-erbB2 receptor tyrosine kinase" Oncogene, 13, 901 (1996).
16. R. Plattner et al., "Loss of oncogenic *ras* expression does not correlate with loss of tumorigenicity in human cells" Proc. Natl. Acad. Sci USA, 93:6665 (1996)
17. D.D. Shoemaker, "Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy", Nature Genet., 14, 450, (1996)

QUALITATIVE ANALYSIS OF GENE NETWORKS

D. THIEFFRY and R. THOMAS

Laboratoire de Génétique des Procaryotes - Université Libre de Bruxelles
Rue des Chevaux, 67, B-1640 Rhode-Saint-Genèse, Belgium
Tel: (32-2) 650 97 26 - Fax: (32-2) 650 97 44
E-mail: {denis,rthomas}@dbm.ulb.ac.be

In this paper, we review the qualitative tools developed by our group for the analysis of regulatory networks. Focusing on the dynamical and biological roles of feedback circuits, this method can be applied in the context of both logical and differential formalisms. This approach already led to several interesting results about the relation between the network structure and the corresponding dynamical properties. In particular, it could be shown that at least one positive regulatory circuit is necessary to generate multistationarity (i.e., alternative states of gene expression), whereas at least one negative circuit is necessary to generate a stable oscillatory behavior. Applications to the analysis of complex gene networks, as well as to the synthesis of regulatory models to account for global expression data are discussed.

1. Introduction

This decade will probably be remembered as the "genome decade". Indeed, almost a dozen of microorganism sequences have already been completed, including mainly bacteria but also *S. cerevisiae*. In addition, many other genomic projects are well on their way, including those dealing with Man, Mouse, *A. thaliana*, *C. elegans*, and *D. melanogaster*. However, there is a long way to go from a complete genomic sequence to a functional understanding of the corresponding organism. Even in the case of *E. coli*, the best characterized free-living organism, the recent completion of the DNA sequence let us with a lot of open questions regarding gene function, regulatory mechanisms, or global integration.

Besides genome sequencing, a series of large scale analyses have been initiated, aiming at uncovering the functional organization of cells. In order to disentangle gene regulatory networks at the level of the whole organism, several groups started systematic global studies of gene expression and DNA-protein interactions in different conditions (1, 39, 40). Clearly, such time or space scale snapshots of gene expression in various conditions will be of great help in the delineation of the main regulatory pathways. As a complement to these experimental approaches, there is an increasing need for efficient theoretical tools and formal frameworks to derive regulatory structures from partial expression data (3, 4, 15, 23).

About three decades ago, several groups independently started to develop qualitative tools for the dynamical analysis of gene regulatory networks (5, 8-10). In this paper, we review the work performed at the Université Libre de Bruxelles, leading to the development of a set of theoretical concepts and formal tools which