THEORETICAL MATHEMATICL BIOLOGY

Edited by

R. T. Hancock and C. K. Zhong

Lanzhou University Press

THEORETICAL MATHEMATICAL BIOLOGY

Edited by

R.L.Hancock and C.K.Zhong

Lanzhou University Press

(甘)新登字第 08 号

理 论 生 物 数 学

韩荣礼　钟承奎　主编

兰州大学出版社出版发行
(兰州大学校内)
甘肃静宁印刷厂印刷
开本: 1850: 1168 毫米 1／32　　印张: 4.125
1994年9月第 1 版　　1994 年 9 月第 1 次印刷
字数: 100 千字　　印数: 1—1000 册
ISBN7-311-00759-3／Q · 19　　定价: 14.00 元

We wish to dedicate this book to our Chincse friends.

RLH and CKZ

Introduction

The following papers in mathematical biology are the outcome of work by the staff during the first year of the newly established Chinese Institute of Theoretical Biology at Lanzhou University in China. Presently, the area of research emphasized at the institute is on control mechanisms of embryonic genes—a topic which is reflected in many of these papers.

The papers are presented for their ideas and are not to be considered as an attempt at any rigorous treatment of the subjects. Furthermore, we make little apology for any radical use of terms and mathematical expressions for we are exploring possibilities—not finalities.

Dr.R.L.Hancock, Director
Dr.C.K.Zhong, Staff Scientist
The Chinese Institute of Theoretical Biology

Acknowledgements

We wish to thank Mr. S. H Ren, Mr. J. S. Wang, Mr.Z.P.Gu, and the Lanzhou University research office for contributions which made this publication possible.

CONTENTS

Vector Algebraic Expressions of Embryonic Gene Activity

R. L. Hancock *V. W. S. Leung*

Abstract

Gene activity pseudo–vector space is developed and embryonic gene activity vectors are considered as a subset of activity vectors. Enhancer elements act as multiplicative scalars. The concept of a spanning set of embryonic gene activity vectors is explored. Certain gene activities could be considered as being controlled by a combination of gene activities from the spanning set of genes. Gene activity operators can be used to describe changes in rates of activity during various phases of the cell cycle and during differentiation. The effects of carcinogenic chemicals can be described as perturbing projection operator functions causing dysdifferentiation. Lastly transformation matrices can act as repressors on vector matrices.

Introduction

The control mechanism for embryonic genes is of great

importance to our understanding of differentiation and dysdifferentiation (carcinogenesis). Carcinogenic chemicals can induce embryonic gene activity such as ethionine–induced alpha–fetoprotein synthesis (Hancock,et. al 1976). A theoretical model has been developed for this induction process (Hancock,1991).

We wish now to develop mathematical expressions that can be used in studying theoretical models of embryonic gene control in a rigorous manner. The following is an approach to the use of vector algebra in such descriptions.

Mathematical Development

Gene Activity Pseudo–vector Space. A vector may be defined as an ordered pair of quantities [components] V(a,t) where a is the amount of gene product and t is time. Therefore, we may use a vector $\vec{g}$ to represent gene activity. Gene activity, in this paper, is defined as the rate of reaction of RNA polymerase using a DNA (gene) template. Note that if $\vec{g} = (a,0)$ then $\vec{g} = (0,0) = \vec{o}$ since any activity at $t = 0$ would be a zero vector and if $\vec{g} = (0,t)$ then $\vec{g} = (0,0) = \vec{o}$ since any activity with a zero amount of product would result in a zero activity vector. Since a and t are real numbers, it follows that $\vec{g}$ is in $R \times R$. Let G denote the set of all gene activity vectors, i. e. $G = \{\vec{g} \mid \vec{g}$ is a gene activity vector. $\}$. We shall show that G with the operations of addition and scalar multiplication has many of the

properties of a vector space.

Let $\vec{g}_1$ and $\vec{g}_2$ be two gene activity vectors in G. Then $\vec{g}_1 + \vec{g}_2$ is in G. For example, if $\vec{g}_1$ is the amount of activity of the gene for alpha-fetoprotein, an embryonic type of albumin, then the total amount of albumin-type gene activity would still be a value in G. For $\vec{g}_1$, $\vec{g}_2$ and $\vec{g}_3$ in G, we also have $(\vec{g}_1 + \vec{g}_2) + \vec{g}_3 = \vec{g}_1 + (\vec{g}_2 + \vec{g}_3)$. Here the associative law of gene activity vectors would apply. There is an $\vec{o}$ vector in G such that for all $\vec{g}$ in G, $\vec{g} + \vec{o} = \vec{o} + \vec{g} = \vec{g}$. Such an $\vec{o}$ genic activity vector would be the activity at zero time or of zero amount of product or both. Corresponding to each $\vec{g}$ in G, the additive inverse vector $-\vec{g}$ exists in G such that $\vec{g} + (-\vec{g}) = (-\vec{g}) + \vec{g} = \vec{o}$, since the gene activity is the product mRNA, any breakdown (e. g. that produced by ribonuclease) of the product could be considered the negative genic activity vector. If $\vec{g}_1$ and $\vec{g}_2$ are in G, then $\vec{g}_1 + \vec{g}_2 = \vec{g}_2 + \vec{g}_1$. If $\vec{g}$ is in G and k is a scalar, then k $\vec{g}$ is in G. Furthermore if $\vec{g}_1$ and $\vec{g}_2$ are in G and k is a scalar, then $k(\vec{g}_1 + \vec{g}_2) = k\vec{g}_1 + k\vec{g}_2$ and if $\vec{g}$ is in G and k_1 and k_2 are scalars, then also for every $\vec{g}$ in G, we have $1 \times \vec{g} = \vec{g}$. where I is called the multiplicative identity.

In general a scalar k can be interpreted biologically as an enchancer[Sassone-Copsi, et, al 1984] or suppressor[Muqlin, et. al 1986]. Enhancers act as multiplicative scalars [Kay, et, al. 1987]. For

example, two copies of a 72 base–pair repeat enhancer element increased the efficiency of activation of a particular gene expression and increasing the elements to four further increased the activation of gene expression proportional to the number of enhancers added[Kuman, et, al, 1986]. However, further additions diminished the activation of gene expression proportional to the number experimentally added. Thus the linearity property is not fully satisfied. For this reason we call the set of gene activity vectors G together with the operations of addition and scalar multiplication a pseudo–vector space of gene activities.

Pseudo–subspace of G. Let E represent the set of all embryonic gene activity vectors. Then E is a subset of G. E is a nonempty subset of G since embryonic genes exist. E satisfies the closure property:i) if $\bar{e}_1$ and $\bar{e}_2$ are in E, then $\bar{e}_1 + \bar{e}_2$ is in E. ii) for $\bar{e}$ in E, $k\bar{e}$ is also in E.

Here the examples of alpha–fetoprotein and fetal hemoglobin gene activities would suffice for $\bar{e}_1$ and $\bar{e}_2$ respectively. Thus the set of all embryonic gene activity vectors form a pseudo–subspace of G.

Spanning Sets of Embryonic Gene Activities. One may express linear combinations of gene activities in the following manner, If the total amount of embryonic gene activity is $\bar{e}$ and if $k\bar{e}$ represents a specific amount of modulated activity, because of some regulator that is proportional to k as a repressor (e. g. methylation of a promoter region by DNA methylase) or activator (e. g. azacytidine–inhibited methylation of the promoter region), then the overall or to-

tal embryonic gene activity at any one state of differentiation may be represented as : $\bar{e} = k1\bar{e}_1 + k2\bar{e}_2 + \cdots + k1\bar{e}_1$, where $\bar{e}_2$ for example would be the alpha–fetoprotein gene activity and $\bar{e}_2$ would denote the fetal hemoglobin gene activity, and in general $\bar{e}_1$ would represent the i^{th} embryonic gene activity for $i=1, 2,...j$. Note that the k_i ,..., k_j, components will vary according to when the gene becomes repressed or derepressed during differentiation until some specific or designated state.

First we will mimic the strict mathematical definition of a spanning set of vectors for a pseudo–vector space. Then we will attempt to derive some biological spanning concepts. The span of a pseudo–vector space of gene activities is defined as follows: let $E^* = \{\bar{e}_1 ,...\bar{e}_n \}$ be a subset of a subset of E. The set E^* is said to span E if every $\bar{e} \in E$ can be written as a linear combination of the elements of E^*. That is , $\forall \bar{e} \in E$ scalars $a_1,...,a_n$ such that $\bar{e} = a_1\bar{e}_1 + \cdots + a_n\bar{e}_n$.

Biologically one could conceive of several approaches to a spanning set of activity gene vectors. For example, a theoretical, yet potentially real process, is the use of a spanning set of embryonic gene vectors described as follows. Let $\{\bar{e}_1 , \bar{e}_2 , \bar{e}_3 \}$be a set of three activity vectors where $\bar{e}_1$ is the gene activity for a specific chromatin protein methylase, $\bar{e}_2$ is the gene activity for a specific chromatin phosphoprotein kinase and $\bar{e}_3$ is the activity for a specific DNA methylase. Let $\bar{e}_1$ be amount

of embryonic gene aetivity for embryonic albumin–like proteins and let $\bar{e}_1$ be the amount of embryonic gene activity for embryonic type hemoglobin proteins. In the former case it is noted that a protein methylase and phosphokinase are required for embryonic albumin genes. For the embryonic hemoglobin activity a specific DNA methylase activity is also necessary for some specific chromatin configurational change in this hypothetical case. Thus we see that a specific set of genes could control, through their various combinations, the chromatin configuration allowing for activity of different kinds of embryonic genes. Therefore this set of genes $\{\bar{e}_1, \bar{e}_2, \bar{e}_3\}$ would be considered as a biological spanning set. We may write $\bar{e}_1 = a_1\bar{e}_1 + a_2\bar{e}_2$ and $\bar{e}_1 = b_1\bar{e}_1 + b_2\bar{e}_2 + b_3\bar{e}_3$. In general such spanning sets could be used for the interpretation and description of specific types of embryonic gene activities.

One may also explore the idea of considering the enhancer regions of structural genes as a spanning set for the pseudo–vector space of gene activity vectors. Note that since enhancer modulated activity is not linear one must again deal with a pseudo–vector space.

Let us expand our discussion further and consider some gene activities that are due to control type genes[Jacob and Monod, 1969]. The rate of synthesis of other proteins from structural type genes have been shown to be governed by control type genes. This is typified by the classical operon model which has genetic elements of a regu-

lator gene, operator gene and a set of structural genes. One may speculate that the control type gene activities may be considered a spanning set. Such a spanning set of gene activities would be a fundamental system which controls other gene activities.

In the mathematical case unit vectors $\bar{u}_1$, $\bar{u}_2$ and $\bar{u}_3$ in R^3, the 3–dimensional space, is an example of a spanning set and combinations of the spanning set along with scalars can produce any vector in R^3.

In a mathematical biological system an analogous system could be as follows. There would be "unit" gene activity vectors of the controlling type genes. Let $\bar{e}_a$, $\bar{e}_b$, $\bar{e}_c$ be three such controlling gene activity vectors of a hypothetical gene system, i. e. a specific set of control and structural genes. Let E_1 be comprised of a series of gene activities that produce three trans–acting protein factors which will regulate any number of embryonie type struetural gene activities. Various protein complexes (transcription complexes) are known to exist at structural gene initiation sites[Beardsey, 1991]. Thus combinations of $\bar{e}_a$, $\bar{e}_b$ and $\bar{e}_c$ would act as a spanning set and "produce" or control any activity vector arising in E_1.

It is recognized that in the mathematical case a spanning set may form the basis of an infinite set of vectors in R^3 whereas in the biological case only a finite set of activity vectors are spanned by the spanning set of control gene activity vectors.

An expansion of these ideas can be devised using the

concept of methylation of specific proteins in chromatin. Such methylations could be critical to the structure of regions of heterochromatin. Let E (sp) be the gene activities that have been coded for particular chromatin protein methylases. These would be secondary control type genes that allow (repress) or disallow (activate) heterochromatization of a series of genes. We may then call E(sp) a spanning set of activity genes. Therefore this spanning set would determine other gene activity vectors via the heterochromatization process.

A particular type of chromatin protein methylase namely histone methylase activity has been shown by one of us (R. L. H.) to be increased in embryonic and neoplastic liver tissues[Turner and Hancock, 1970] Furthermore, it is also increased after the administration of a hepato carcinogen[Hancock 1978]. This introduces a correlation between chromatin methylation and perturbation of gene expression.

Gene Activity Operators. (I)The projection operator P. Let $\bar{e}_1$, $\bar{e}_2$,..., $\bar{e}_n$ be a set of gene activity vectors in an n–dimensional embryonic gene activity pseudo–vector space R^m. Associate with the vector.

$\bar{e}^n = \sum_{k=1}^{n} a_k \bar{e}_k$ the new vector:

$\bar{e}^m = P\bar{e}^n = \sum_{k=1}^{m} a_k \bar{e}_k$ with $m < n$. P is defined to be a projection operator.

The operator P projects a given vector $\bar{e}_n$ in R^n into the subspace R^m spanned by the vectors: $\bar{e}_1$, $\bar{e}_2$,...,$\bar{e}_m$. This represents a process by which new embryonic gene activities are formed. The linear combination of $a_1\bar{e}_1 + a_2\bar{e}_2 + \cdots + a_m\bar{e}_m$ represents the formation of new gene activity $\bar{e}_m$ which would occur during differentation. We interpret the biological operator P as a differentiation function that represents a gene activity induction process.

One could speculate that carcinogenic chemicals induce embryonic gene activities by causing anomalous functioning of some of the projection operators. Once the carcinogen has perturbed one or more gene activities and if these activities are members of a fundamental spanning set of activites. then biologically normal projected activity vectors $\{\bar{e}_{n1}$, $\bar{e}_{n2}$,..., $\bar{e}_{nk}\}$ would be placed onto an anomalous set of spanning vectors generating new vector activties $\{\bar{e}^1_{n1}$, $\bar{e}^1_{n2}$,..., $\bar{e}^1_{nk}\}$ not encountered during differentiation and thus creating a state of dysdifferentiation.

The following illustrates this idea diagramatically: