Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 409 (2005) 70–78
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Bipartite and tripartite systems and matrices from genetic control research☆

Clark Jeffries *

Department of Psychiatry, School of Pharmacy, and Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States

Received 16 November 2004; accepted 13 March 2005
Available online 24 August 2005
Submitted by S. Kirkland

Abstract

Discovering the organizational principles of genetic expression has recently become an arena of substantial investigative effort and modeling challenges. Laboratory findings of geneticists can be couched in terms of classes of stoichiometric networks that through stability analysis lead to classes of matrix patterns. In particular, targets, blocks, and decoys are variables in genetic systems with relationships that can be described by bipartite and tripartite graphs. Related dynamical systems will exhibit stability if a mixture of qualitative and quantitative criteria is applied. Analyses of the models suggest limits of total induction rates of blocks and decoys relative to other rates.

AMS classification: 92C45; 80A30

Keywords: Gene regulation; Combinatorial control

☆ USP#SMH01905-01 North Carolina Foundation of Hope, Anonymous donor administered through the San Francisco Foundation.
* Tel.: +1 919 260 6576.
E-mail address: clark_jeffries@med.unc.edu

0024-3795/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2005.03.039

1. Introduction

In recent years and especially in recent months, some basic notions of the science of genetics have been modified by discoveries regarding the regulation of gene expression. One discovery [2] is that "a few molecules" of small, double-stranded RNA molecules induced synthetically into a cell (in a worm) can escape notice of the immune system, replicate, and then alter gene expression to cause macroscopic changes. Another is the seemingly essential role of noncoding RNA transcripts in control of genetic expression, specifically acting as decoys relative to agents that block translation of similar mRNAs [3]. These and related discoveries about RNA molecules that control genetic expression but are not used directly as protein templates (as messenger RNA (mRNA)) have caused some geneticists to consider supplementing the "central dogma" (that all biologically important information in chromosomes is used as protein templates).

Conventional transcriptional control includes feedback mechanisms from cellular metabolite levels that regulate in several known ways the primary rate of transcription of a gene. However, some large genes take many minutes or even hours to be transcribed into RNA. Their production might be "pipelined" with several transcripts in various stages of completion at any time. In order to respond quickly (say, in seconds) to changing protein demands, cellular machinery might more quickly produce small blocking RNA molecules that at various times prevent some or many large transcripts from proceeding to translation to protein. This hybrid system might be more responsive than conventional regulation. At any rate, many geneticists now feel that paradigms for conventional regulation should be supplemented to some degree with regulation by noncoding RNA.

As shown below, stoichiometric models of genetic control systems lead to interesting bipartite and tripartite matrix patterns as linear approximation matrices. It is known that some mRNA molecules (called targets) can be precluded from translation by binding with small subsequences with one type of several types of small, blocking molecules (called herein blocks). Some blocks such as so-called microRNAs can prevent translation of more than one target and some targets might be blocked by more than one microRNA [8]. According to a bioinformatics survey [9], the $\sim$220 known human microRNAs might have 13,000 target sequences in 5300 genes.

The human genome also transcribes pseudogenes (RNA molecules much like mRNA but lacking essential components to be translated into protein). Some pseudogenes act as decoys for blocking actions. Thus, there are three types of components (at least) in genetic control systems: targets, blocks, and decoys. As vertices in a graph, targets are joined by edges to blocks and blocks are joined to decoys; no other types of edges are allowed herein. The graph of actual targets, blocks, and decoys for the human genome must be quite large. Likely, it has some type of hierarchical organization, the focus presently of much research (e.g., [6]).

The virtues of sign-stable or qualitatively stable systems as candidates for high-level paradigms in ecology and perhaps other disciplines have often been stated [7].

However, qualitative stability of linear systems is a restrictive condition; it would chafe modelers of complex phenomena such as genetic control systems. More appealing might be nonlinear models that have a few, simple quantitative conditions that neatly imply stability; otherwise the nonlinear "near-qualitative" models are still stable (many or most model coefficients can be of arbitrary magnitude without affecting the qualitative existence of a stable constant trajectory). The further elucidation of such systems would seem to be a major challenge to applied mathematics. This paper is about a class of such nonlinear models that arise in genetic control. Modeling the above and other putative components of the control system for gene expression is a goal that has the attention of many researchers. The research is motivated by the apparent potential for new understanding and perhaps new treatments of complex diseases [10].

2. Genetic control model

The model considered herein includes a number $x > 0$ of targets $T_1, T_2, \ldots, T_x$, a number $m > 0$ of blocks $B_1, B_2, \ldots, B_m$, and a number $n\ \mu\ 0$ of decoys $D_1, D_2, \ldots, D_n$. These substances (or their precursors) are all created by transcription in the nucleus of a cell. After some processing, they are exported into the cytosol to interact as in a chemical reaction network. The proteins expressed by the system are functions of target levels (reacting with ribosomes), but protein levels are not directly used in the present analysis. Thus, a model with targets, blocks, and decoys might be written as follows:

$$\frac{\mathrm{d}T_a}{\mathrm{d}t} = \tau_a - \sigma_a T_a - T_a \left[\sum_i \rho_{ai} B_i\right], \quad a = 1, 2, \ldots, x, \tag{1}$$

$$\frac{\mathrm{d}B_i}{\mathrm{d}t} = \beta_i - B_i \left[\sum_a \rho_{ai} T_a\right] - B_i \left[\sum_j \mu_{ij} D_j\right], \quad i = 1, 2, \ldots, m,$$

$$\frac{\mathrm{d}D_j}{\mathrm{d}t} = \delta_j - D_j \left[\sum_i \mu_{ij} B_i\right], \quad j = 1, 2, \ldots, n.$$

Nine types of positive rate coefficients are: τ_a, the creation of target T_a; σ_a, the removal of T_a in self-regulation mechanisms; ρ_{ai}, the blocking reaction of T_a and B_i, β_i, the creation of block B_i; and δ_j, the creation of decoy D_j. Thus the rate of reaction of T_a and B_i (zero or positive) is ρ_{ai}, a nonnegative matrix with at least one positive entry in every row and at least one positive entry in every column. Thus we assume that every target reacts with at least one block and every block reacts with at least one target. The rate of reaction of B_i and D_j (also nonnegative) is μ_{ij}. We also assume that any decoy must affect at least one block, so every column of the *n-by-m*

nonnegative matrix μ_{ij} has at least one positive entry. However, some block might react with no decoys, so an all-zero row in μ_{ij} is possible.

To organize the relationships and, as shown below, to express stability conditions, a combinatorial description of relationships among targets, blocks, and decoys is needed; the format we use is a tripartite graph called herein a *TBD graph*. In a *TBD* graph, targets appear as vertices of a first type, blocks as vertices of a second type, and decoys as vertices of a third type; edges denote processes. For example, suppose targets $[T_1, T_2, T_3]$ are blocked by $[B_1, B_2, B_3]$, $[B_2, B_4]$, and $[B_4]$, respectively. Suppose blocks $[B_1, B_2, B_3, B_4]$ react with decoy sets [*empty*], $[D_1, D_2, D_3]$, $[D_3]$, and $[D_4]$, respectively. The associated tripartite graph is shown in Fig. 1.

Fig. 1 depicts a small example of a connected (always assumed) *TBD* graph (every vertex can be reached by every other vertex by following a path of edges).

We will also need to consider the bipartite graph called herein a *BD graph* obtained by deleting all the T vertices and their edges from the *TBD* graph from (1). The *BD* graph generally is not connected. In Fig. 1, B_1 is not connected to any D vertex and so is a trivial maximal connected subgraph of the *BD* graph; likewise $[B_4, D_4]$ are in a maximal connected subgraph of the *BD* graph.

To exhibit a stable equilibrium, the high-dimensional model turns out to have algebraic requirements on the sums of certain induction rates associated with the entire *TBD* graph and the maximal connected components of the *BD* graph. The speed at which the system approaches stable equilibrium depends generally on all model coefficients, but the existence of stable equilibrium depends only on such sums of induction rates. This qualitative stability seems to be a distinguishing property among high-dimensional nonlinear models [4].

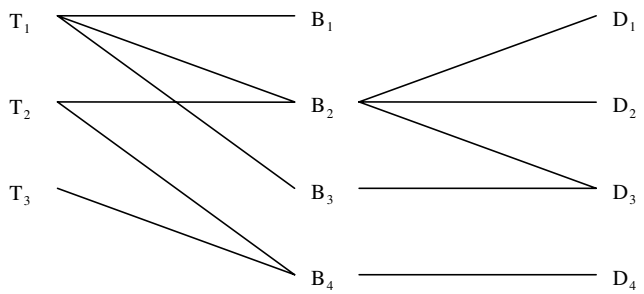

Fig. 1. A tripartite graph showing three types of vertices: three mRNA targets, four small-molecule RNA blocks, and four RNA decoys that have subsequences that are similar to those of targets. Edges denote processes. In general, there can be any positive number of T or B vertices; if D vertices occur, they can be of any number. For the existence of a constant state with all variables positive (feasible), it turns out that the total input rates of B type must exceed that for D type in a certain subgraph. Also, the difference of total B induction and total D induction must be less than the induction sum for the T types. Any such feasible constant state is automatically at least locally stable. The process rates between variables are otherwise of arbitrary magnitude, so this model exhibits a kind of qualitative stability.

Theorem 1 (Existence of constant feasible state). *A necessary condition for the existence of a feasible* (*all components positive*) *constant trajectory for* (1) *is* $\sum_a \tau_a > \sum_i \beta_i - \sum_j \delta_j$ *where the sums are taken over the indices of vertices in the TBD graph. Another necessary condition is that sums taken over the B and D subsets in each maximal connected subgraph of the BD graph fulfill* $\sum_i \beta_i > \sum_j \delta_j$.

Proof. Suppose a feasible constant trajectory exists; let us denote it as $(T_1, \ldots, T_x, B_1, \ldots, B_m, D_1, \ldots, D_n)$. The left side of each equation in (1) is zero, so the sum of all the right sides of (1) must also be zero. It follows that $\sum_a \tau_a - \sum_i \beta_i + \sum_j \delta_j = \sum_a \sigma_a T_a$ and consequently $\sum_a \tau_a - \sum_i \beta_i + \sum_j \delta_j$ must be positive.

Furthermore, for the B and D vertices in a maximal connected subgraph of the BD graph, let us denote the associated submatrix with i, j entry by $\mu_{ij} B_i D_j$. The ith row sum must be $\beta_i - \sum_a \rho_{ai} T_a B_i < \beta_i$ and the jth column sum must be δ_j. Summing all entries is independent of order of summation, hence $\sum_i \beta_i > \sum_j \delta_j$. □

In the special case that only one target T_1 is present and no decoys are in the system, the unique constant state can be readily calculated: $T_1 = (\tau_1 - \sum_i \beta_i)/\sigma_1$ and $B_i = \beta_i/(\rho_{1i} T_1)$. In this case, a necessary and sufficient condition for the existence of a constant state with all components positive applies just to the induction rates of target and blocks, namely, $\tau_1 > \sum_i \beta_i$.

Regarding stability, the negative of the linear approximation matrix $-L$ of (1) at a feasible constant state is

$$-L = \left(\begin{array}{c|c|c} \sigma_1 + \sum_i \rho_{1i} B_i \;\cdots\; 0 & T_1\rho_{11} \;\cdots\; T_1\rho_{1m} & 0 \;\cdots\; 0 \\ 0 \;\cdots\; \sigma_x + \sum_i \rho_{xi} B_i & T_x\rho_{x1} \;\cdots\; T_x\rho_{xm} & 0 \;\cdots\; 0 \\ \hline \rho_{11}B_1 \;\cdots\; \rho_{x1}B_1 & \sum_a \rho_{a1}T_a + \sum_j \mu_{1j}D_j \;\cdots\; 0 & B_1\mu_{11} \;\cdots\; B_1\mu_{1n} \\ \rho_{1m}B_m \;\cdots\; \rho_{xm}B_m & 0 \;\cdots\; \sum_a \rho_{am}T_a + \sum_j \mu_{mj}D_j & B_m\mu_{m1} \;\cdots\; B_m\mu_{mn} \\ \hline 0 \;\cdots\; 0 & \mu_{11}D_1 \;\cdots\; \mu_{m1}D_1 & \sum_i \mu_{i1}B_i \;\cdots\; 0 \\ 0 \;\cdots\; 0 & \mu_{1n}D_n \;\cdots\; \mu_{mn}D_n & 0 \;\cdots\; \sum_i \mu_{in}B_i \end{array}\right). \tag{2}$$

Note that in $-L$ (considered as nine subblocks) subblocks 13 and 31 are all 0; subblocks 11, 22, 33 are in diagonal form; the other subblocks in general are full, rectangular matrices. In (2), it is assumed that each $\sigma_a > 0$; for each j there is an i such that $\mu_{ij} > 0$; for each i there is an a such that $\rho_{ai} > 0$; and for each a there is an i such that $\rho_{ai} > 0$. Also, (1) and the assumptions $\tau_a, \beta_i, \delta_j > 0$ imply $T_a, B_i, D_j > 0$.

Theorem 2 (Stability of a constant feasible state). *The system* (1) *is stable at any feasible constant trajectory.*

Proof. It will suffice to show the real part of every eigenvalue of $-L$ in (2) is positive. In the matrix $-L$ all entries are nonnegative. For each column, we can compare the sum of off-diagonal entries with the diagonal. The diagonal entry is greater for the columns intersecting subblock 11 and equal for the rest. According

to the column version of Gershgorin's theorem [5] all the eigenvalues of $-L$ are in certain Gershgorin disks in the complex plane. The centers of the Gershgorin disks are on the real axis and have as centers the diagonal entries of $-L$. The radii of the discs for $-L$ are the off-diagonal column sums. It follows that all the points in the union of all the disks lie in the positive half-plane with one possible exception: the origin. The origin amounts to the possibility of 0 as an eigenvalue for $-L$. Therefore the linear approximation matrix L would be stable, implying local stability of the feasible state, provided 0 were provably not an eigenvalue of $-L$.

Let A be the diagonal matrix with diagonal entries $(T_1, \ldots, T_x, B_1, \ldots, B_m, D_1, \ldots, D_n)$. Define a symmetric matrix $M = -LA$; clearly 0 is an eigenvalue of $-L$ if and only if zero is an eigenvalue for M.

In hope of deriving a contradiction, suppose 0 is an eigenvalue of $-L$ and so let z be a nonzero $(x + m + n)$-vector that solves $Mz = 0$. Denote the first x components of z by $u = u_1, \ldots, u_x$, the next m components by $v = v_1, \ldots, v_m$, and the last n components by $w = w_1, \ldots, w_n$.

Let us partition the relationships of u, v, w as follows. Case 1: One of the u components of z is at least a large in magnitude as any other component of z. Case 2: Failing the previous case, some component of v is greater in magnitude than any u component and at least as large as the magnitude of any w component. Case 3: Failing the previous cases, it must be that some component of w is greater in magnitude than any component of u or v.

Case 1. Some component of u, namely, without loss of generality $z_1 = u_1 > 0$, is at least as great in magnitude as any other entry in z. The number 1 row equation of $Mz = 0$ yields $T_1\big(\sigma_1 + \sum_i \rho_{1i} B_i\big)u_1 + \sum_i T_1 \rho_{1i} B_i v_i = 0$. Therefore, $T_1\big(\sigma_1 + \sum_i \rho_{1i} B_i\big)u_1 = \big|\sum_i T_1 \rho_{1i} B_i v_i\big|$ where $|\cdot|$ denotes absolute value. Thus

$$\left(\sigma_1 + \sum_i \rho_{1i} B_i\right) u_1 \leqslant \sum_i \rho_{1i} B_i |v_i|. \tag{3}$$

Since every $|v_i| \leqslant u_1$ and since σ_1 is positive, (3) is impossible to solve.

Case 2. Some component of v, namely, $v_1 > 0$, is larger than the magnitude of any component of u and at least as large as any component of w. The number $x + 1$ row equation of $Mz = 0$ yields $\sum_a \rho_{a1} T_a B_1 u_a + \sum_a \rho_{a1} T_a B_1 v_1 + \sum_j \mu_{1j} D_j B_1 v_1 + \sum_j \mu_{1j} B_1 D_j w_j = 0$. Therefore, $\big(\sum_a \rho_{a1} T_a B_1 + \sum_j \mu_{1j} D_j B_1\big)v_1 = \big|\sum_a \rho_{a1} T_a B_1 u_a + \sum_j \mu_{1j} B_1 D_j w_j\big|$. Thus

$$\left(\sum_a \rho_{a1} T_a + \sum_j \mu_{1j} D_j\right) v_1 \leqslant \sum_a \rho_{a1} T_a |u_a| + \sum_j \mu_{1j} D_j |w_j|. \tag{4}$$

Since every $|u_a| < v_1$, at least one of $\{\rho_{a1}\}$ is positive, and every $|w_j| \leqslant v_1$, it follows that (4) is impossible to solve.

Case 3. Some component of w, namely, $w_1 > 0$, is larger than the magnitude of any component of v. The number $x + m + 1$ row equation of $Mz = 0$ yields $\sum_i \mu_{i1} D_1 B_i v_i + \sum_i \mu_{i1} D_1 B_i w_1 = 0$. Therefore $\sum_i \mu_{i1} D_1 B_i w_1 = \left| \sum_i \mu_{i1} D_1 B_i v_i \right|$ and so

$$\sum_i \mu_{i1} B_i w_1 \leqslant \sum_i \mu_{i1} B_i |v_i|. \tag{5}$$

Since every $|v_i| < w_1$ and since at least one of $\{\mu_{i1}\}$ is positive, (5) is impossible to solve. □

(A referee for this paper has kindly shown me that Theorem 2 could be proven in an alternative manner using theorems pertaining to M-matrices and Z-matrices from a classic of linear algebra [1]. In particular, if R is defined to be formed from the identity matrix of the size of $-L$ except having the subblock 22 containing only -1 on its diagonal, then it can be shown that $R(-L^{\mathrm{T}})R^{-1}$ is a Z-matrix. From further results in the cited text, especially page 136, condition (L_{33}), it can be shown that this matrix is a nonsingular M-matrix, from which the conclusion of Theorem 2 follows. I am indebted to the referee for an exceedingly careful review of the manuscript and also for demonstrating the power of advanced theorems in the cited text.)

I conjecture that the inequalities in Theorem 1 are actually necessary and sufficient for (1) to have a constant, unique, feasible, global attactor trajectory. Needed is some kind of global Lyapunov-like analysis, but that does not seem to be available from routine thinking, at least by me. The description of additional nonlinear systems with global, near-qualitative stability would seem to be a worthy mathematical goal.

The biological relevance of Theorems 1 and 2 is the following. First, it is important to realize that (1) in conformance with the above inequalities does not include protein dynamics or feedback; rather, such a model (1) is a stable component of a complete system. Each target mRNA T_a is spontaneously removed from the cell by some auto-degradation mechanism at a rate proportional to T_a itself. Still, none of the m blocks or n decoys has self-regulation. Self-regulation of B or D terms could be added if biologically justified, but the system is already inherently and qualitatively stable. A model using the above could be stable in reaction to perturbations such as when the rate of export from a cell of a protein rapidly increases due to demand for it from the rest of the organism. In summary, the natural stability of (1) in conformance with the conditions would act to maintain target levels by automatic changes in block and decoy levels. Since the numbers (m and n) of the blocks and decoys in the above analysis are arbitrary, one gene with one mRNAs may actually be controlled by any number of blocks; furthermore, that control might be tuned by any number of decoys.

The key point is that the block transcription rate depends upon the current target mRNA level, not any past level of the protein product P. (In fact, using blocks with transcript rates determined by sufficiently old P values could destabilize the model.) This is much in contrast with conventional regulation of target transcription rates by pathways involving cellular metabolites, pathways that might be slow, complex,

and, due to time delays, inconsistent with stability in some cases. This might explain why nature might use fast transcription of short blocking molecules to control steady, pipelined, and seemingly rather wasteful production of long mRNAs.

The emphasis here has been on stability of feasible states, but in nature, the actual emphasis might actually be on adjusting the sums of block or decoy production rates to violate the conditions of Theorem 1 and thereby drive protein production to extremes (full rate or nearly nothing). The interesting point is that the induction sums matter, not necessarily the induction rate of any one block or decoy.

3. Implications for genetic control

The model above indicates that in certain stoichiometric networks targets, blocks, and decoys can exhibit a kind of qualitative stability, at least in local linear approximations. In addition, analysis of the model type suggests properties of total induction rates of blocks and decoys, as well as strategies for avoiding instability due to transcription or translation time delays. Certain induction rates will drive protein production to a stable, positive value; others will cause mRNA levels to drop, eventually eliminating protein production as well; still others would drive blocks to low levels, maximizing protein production.

Nature created the gene regulation system over billions of years with random mutation and selection. It might be impossible to discern the existence or importance of every significant control component because the system is embedded in a huge, highly redundant network with apparently many obsolete, nonfunctional parts. But through qualitative analysis inspired by the pioneering work of Professor van den Driessche, scientists might someday be led to understand and treat the crucial control components that are responsible for disease conditions. Her accomplishments and insights in qualitative analysis of many biological systems and many other arenas amount to a wonderful gift to humankind.

References

[1] A. Berman, R. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Classics in Applied Mathematics, vol. 9, SIAM, Philadelphia, 1994.
[2] A. Fire, S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, C.C. Mello, Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, Nature 391 (1998) 806–811.
[3] S. Hirotsune, N. Yoshida, A. Chen, L. Garrett, F. Sugiyama, S. Takahashi, K. Yagami, A. Wynshaw-Boris, A. Yoshiki, An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene, Nature 423 (2003) 91–96.
[4] M. Hirsch, S. Smale, P. Devaney, Differential Equations, Dynamical Systems, and an Introduction to Chaos, Elsevier Academic Press, Burlington, MA, 2004.
[5] R. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, England, 1985.

[6] K. Irizarry, B. Merriman, M. Bahamonde, M.-L. Wong, J. Licinio, The evolution of signaling complexity suggests a mechanism for reducing the genomic search space in human association studies, Mol. Psychiatry 10 (2005) 14–26.
[7] C. Jeffries, V. Klee, P. van den Driessche, When is a matrix sign stable? Canad. J. Math. 29 (1977) 315–326.
[8] L. Lee, N. Lau, P. Garrett-Engele, A. Grimson, J. Schelter, J. Castle, D. Bartel, P. Linsley, J. Johnson, Microarray analysis shows that some microRNAs downregulate large number of target mRNAs, Nature 433 (2005) 769–773.
[9] B. Lewis, C. Burge, D. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
[10] D.O. Perkins, C. Jeffries, P. Sullivan, Expanding the 'central dogman': the regulatory role of non-protein coding genes and implications for the genetic liability to schizophrenia, Mol. Psychiatry 10 (2005) 69–78.