

Patterns of Regulation from mRNA and Protein Time Series

Lingchong You and John Yin¹

Department of Chemical Engineering, University of Wisconsin–Madison, 1415 Engineering Drive, Madison, Wisconsin 53706-1691

Received June 21, 1999; accepted November 2, 1999; published online June 28, 2000

The rapid advance of genome sequencing projects challenges biologists to assign physiological roles to thousands of unknown gene products. We suggest here that regulatory functions and protein–protein interactions involving specific products may be inferred from the trajectories over time of their mRNA and free protein levels within the cell. The level of a protein in the cytoplasm is governed not only by the level of its mRNA and the rate of translation, but also by the protein's folding efficiency, its biochemical modification, its complexation with other components, its degradation, and its transport from the cytoplasmic space. All these co- and post translational events cause the concentration of the protein to deviate from the level that would result if we only accounted for translation of its mRNA. The dynamics of such deviations can create patterns that reflect regulatory functions. Moreover, correlations among deviations highlight protein pairs involved in potential protein–protein interactions. We explore and illustrate these ideas here using a genetically structured simulation for the intracellular growth of bacteriophage T7. © 2000 Academic Press

Key Words: data mining; regulation; protein–protein interactions.

INTRODUCTION

The rapid growth of genomic databases and development of technologies for global monitoring of mRNA (Velculescu *et al.*, 1995; Schena *et al.*, 1995) and protein levels (O'Farrell, 1975; Eckerskorn *et al.*, 1992; Henzel *et al.*, 1993) are creating a need for efficient data-mining methods. In the past few years, reverse engineering methods have been proposed to infer the structure of chemical reaction networks near a steady state (Arkin and Ross, 1995; Arkin *et al.*, 1997) or, more abstractly, the architecture of Boolean networks defined by simple logic rules (Liang *et al.*, 1998; Akutsu *et al.*, 1999). However, little effort has been directed toward inferring relationships among elements of real genetic networks. To address this need, we suggest here a framework for organizing dynamic mRNA and protein data, with the aim of identifying characteristic patterns of

function and potential protein–protein interactions. Given initial information about mRNA levels, and the processing rates and distribution of the ribosomes and activated tRNAs that constitute the translation resources, one can estimate how levels of corresponding proteins will change. Other factors, in addition to translation, will influence the actual protein levels observed: processes of protein modification or degradation, protein transport to or from the cytoplasm, regulatory processes, and the formation of protein–protein complexes. When coupled with translation, different modes of protein appearance or depletion will produce patterns of expression that may reflect aspects of the protein's function. To explore this notion, we employ a genetically based simulation for the growth of a bacteriophage to generate dynamic mRNA and protein data. Then we analyze these model data using a simple algorithm and attempt to identify patterns of regulation and infer protein–protein interactions. Finally, we evaluate the algorithm by comparing its results with the known mechanisms in the simulation.

In an effort to understand how the overall growth rate of a virus depends on the rates of the constituent processes that define its development, we have developed a global kinetic simulation for the intracellular growth of bacteriophage T7, a virus that infects *Escherichia coli* (Endy *et al.*, 1997; Endy, 1997). The 39,937-base-pair double-stranded linear DNA genome of T7 consists of 56 genes coding for 59 known proteins. These genes are grouped into three classes based on their function and position in the wild-type genome. Class I genes moderate the transition in metabolism from host to phage, class II genes are generally responsible for T7 DNA replication, and class III genes code for proteins responsible for phage particle formation, maturation, and host lysis (Studier and Dunn, 1983; Dunn and Studier, 1983). An overview of the T7 growth cycle is shown in Fig. 1. The cycle is initiated when the phage binds to its host, a process that is followed by the gradual translocation of the T7 genome into the cell. Transcription of class I genes (Fig. 1a) is catalyzed by the host RNA polymerase (EcRNAP), which recognizes three major promoters positioned near the entering end of the T7 DNA. Once

¹ To whom correspondence should be addressed. Fax: 608-262-5434. E-mail: yin@enr.wisc.edu.

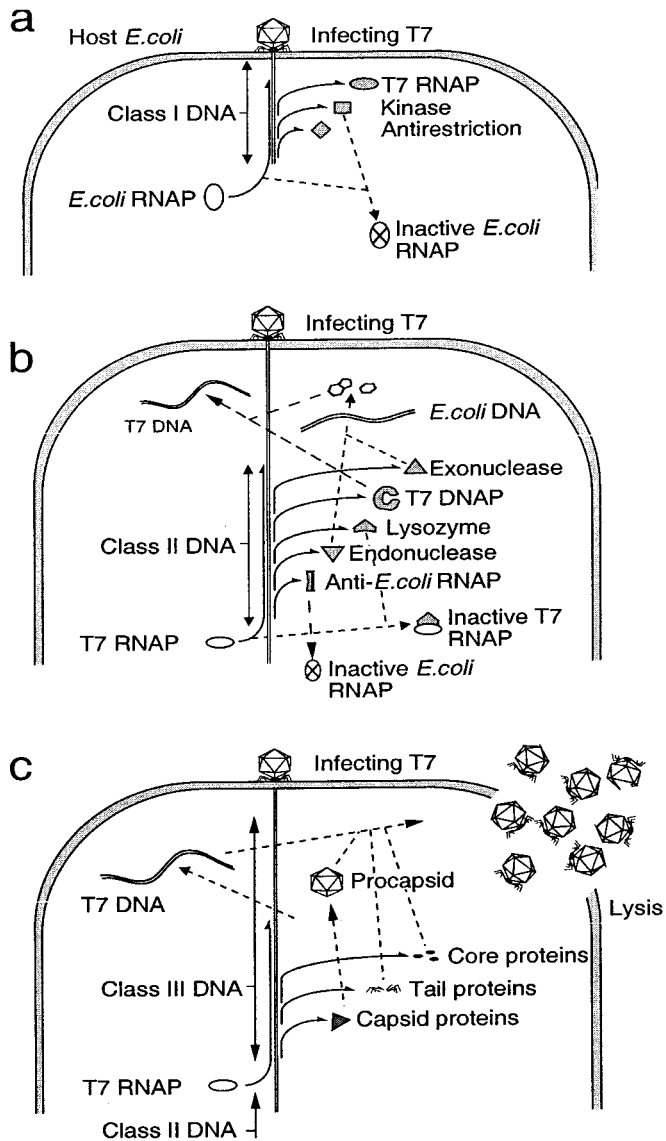


FIG. 1. Intracellular growth cycle of phage T7. The solid lines with half arrows indicate transcriptions and translations, the dashed lines denote reactions, and the solid lines with full arrows indicate the three classes of T7 DNA. (a) Infection initiation, class I gene expression. (b) Class II gene expression, phage DNA replication. (c) Class III gene expression, procapsid assembly, phage maturation, and lysis. RNAP, RNA polymerase; DNAP, DNA polymerase.

translated T7 RNAP (gp1) transcribes the class II (Fig. 1b) and class III (Fig. 1c) genes. Three negative-feedback loops that emerge from the transcription of class I and II genes are noteworthy: gp0.7, a protein kinase, inactivates the *E. coli* RNAP responsible for its own transcription (Fig. 1a); gp2, an *E. coli* RNAP inhibitor, inactivates the *E. coli* RNAP which is responsible for the transcription of the T7 RNAP that it needs for its own transcription (Fig. 1b); and the T7

lysozyme (gp3.5) inactivates the T7 RNAP responsible for its own transcription (Fig. 1b). During the transcription of class III genes T7 procapsids are assembled from the major capsid protein (gp10A), with the help of head assembly protein (gp9). Then, the newly synthesized T7 DNAs, supplemented with additional structural proteins, are packaged into the procapsids to form progeny phage particles (Fig. 1c), which are released by lysis of the host cell. The simulation incorporates the biological mechanisms and rates for T7 genome translocation, mRNA and protein synthesis, DNA replication, procapsid assembly, and phage particle packaging and predicts as output the appearance of approximately 100 phage progeny after 30 min, consistent with experimental observations. It also provides as a by-product the concentration-versus-time trajectories of all T7 mRNAs and proteins (Fig. 2), which we employ in this work as raw data to explore new modes of inference.

COMPUTER SIMULATION

The T7 growth cycle is simulated with a system of coupled ordinary differential equations solved numerically using a fourth-order Runge–Kutta algorithm (Endy *et al.*, 1997). With a 0.032-s time step, the computation time for a typical T7+ growth cycle of 30 min approximately 8 s using a 175-MHz SGI OCTANE workstation. The simulation is stopped by introducing a cutoff time set at 30 min. The mRNA and protein concentrations are sampled every 50 time steps (0.16 s) over the entire simulation. We implement the published simulation (Endy, 1997) with the following modifications:

1. The allocation of *E. coli* RNAP and T7 RNAP to their promoters is based on the promoter strength and is independent of transcript length. In addition, the minor *E. coli* RNAP promoters (B, C, and E) are activated at 5% activity of the major promoters (A1, A2, and A3), based on *in vitro* data (Dayton *et al.*, 1984).
2. A more mechanistic model for T7 DNA replication is adopted. Here, we assume that the DNA replication rate is determined by the number of the limiting species of available replication forks and replication complexes. The number of replication forks is calculated by assuming two replication forks per T7 genome. The number of available replication complexes is determined from the gp4A-to-gp5 replisome stoichiometry of six-to-one (Kusakabe *et al.*, 1998).
3. The implementation of the model has been ported to C++, an object-oriented programming language.

METHODOLOGY

Protein Rate

We define the overall rate change in protein concentration, or the protein rate, as follows:

$$v_i(t) \equiv \frac{dP_i(t)}{dt} = v_{T_i}(t) - v_{D_i}(t), \quad (1)$$

where v_i is the protein rate for protein i , P_i is the concentration of protein i , v_{T_i} is the rate of change in protein i concentration due to translation, and v_{D_i} is the protein i depletion rate. All these variables are the functions of time (t). v_i is determined from levels of free protein taken over time. We assume the protein translation rate, v_{T_i} , is proportional to the concentration of mRNA:

$$v_{T_i} = k_{T_i} R_i, \quad (2)$$

where R_i is the concentration of mRNA encoding protein i and k_{T_i} is the translation rate coefficient for protein i . We assume k_{T_i} is a constant for all proteins that has been measured or can be estimated. For simplicity the index i is dropped in subsequent analyses.

Clearly, in the absence of any depletion effects v_i plotted versus R_i yields a line through the origin with slope k_T . In the presence of depletion effects trajectories on this plot may deviate from the linear behavior in revealing ways. In this work the protein rate is approximated by taking the first-order finite difference of protein concentrations provided by the simulation.

Dynamic Deviation Factor

To quantitatively investigate and compare the deviations of different proteins we introduce a dimensionless parameter, the dynamic deviation factor (DDF), for protein i :

$$D_i(t) \equiv \frac{v_i(t) - v_{T_i}(t)}{v_{T_i}(t)} = -\frac{v_{D_i}(t)}{v_{T_i}(t)}. \quad (3)$$

Further, we define a time-averaged DDF as a measure of the overall deviation of the protein rate from the translation rate:

$$\overline{D}_i = \frac{1}{N} \sum_{k=1}^{k=N} D_{ik}, \quad (4)$$

where D_{ik} is the DDF of protein i at the k th time point of a discretized time course that spans N time points.

Protein Correlation Coefficient

If pairs of proteins associate, their DDFs should be highly correlated. We assess potential pairwise associations by defining a protein correlation coefficient (PCC):

$$C_{ij} \equiv \frac{\sum_{k=1}^{k=N} D_{ik} D_{jk}}{\sqrt{(\sum_{k=1}^N D_{ik}^2)(\sum_{k=1}^N D_{jk}^2)}}, \quad -1 \leq C_{ij} \leq 1. \quad (5)$$

The larger C_{ij} is, the more *likely* two proteins are associating with each other. The extreme case $C_{ij} = 1$ means that the protein rates of the two proteins deviate proportionally from the translation rate in the same direction, i.e., $D_{ik} = \lambda D_{jk}$ for $k = 1, \dots, N$, where λ is a positive constant. A protein correlation matrix (PCM) can be constructed by calculating the pairwise PCCs for all the proteins of interest.

RESULTS

The time series of mRNA and free protein concentrations generated by the simulation for different T7 gene products shown in Fig. 2 reflect little of the diverse enzymatic, structural, and regulatory roles the proteins play during T7 development. Only T7 RNAP (gp1) provides a hint of its regulated activity through its unusual trajectory relative to other proteins (Fig. 2b).

Plotting protein rate versus mRNA concentration for the T7 gene products reveals a diversity of trajectories, as shown in Fig. 3a. For clarity, a representative subset of the total trajectories is shown. A time course is implicit for each trajectory, with an initial condition at the origin where neither mRNA nor protein has been synthesized. All trajectories coincide with or lie below the reference line (labeled “linear” corresponding to translation without any depletion).

The trajectories in Fig. 3a may be coarsely classified as linear or nonlinear. The simplest mechanistic rationale for the linear behavior of gp19, which coincides with the reference line, is that it is generated by translation and lacks any depletion effects. Other gene products, gp0.3, gp2.5, gp4A, gp5, gp6, gp7, gp17.5, and gp18, exhibited similar behavior (not shown).

The nonlinear trajectories may be subdivided into two groups based on their shapes. Trajectories for gp9 and gp10A follow the reference line early and deviate later (Fig. 3a). Other gene products, gp8, gp11, gp12, gp13, gp15,

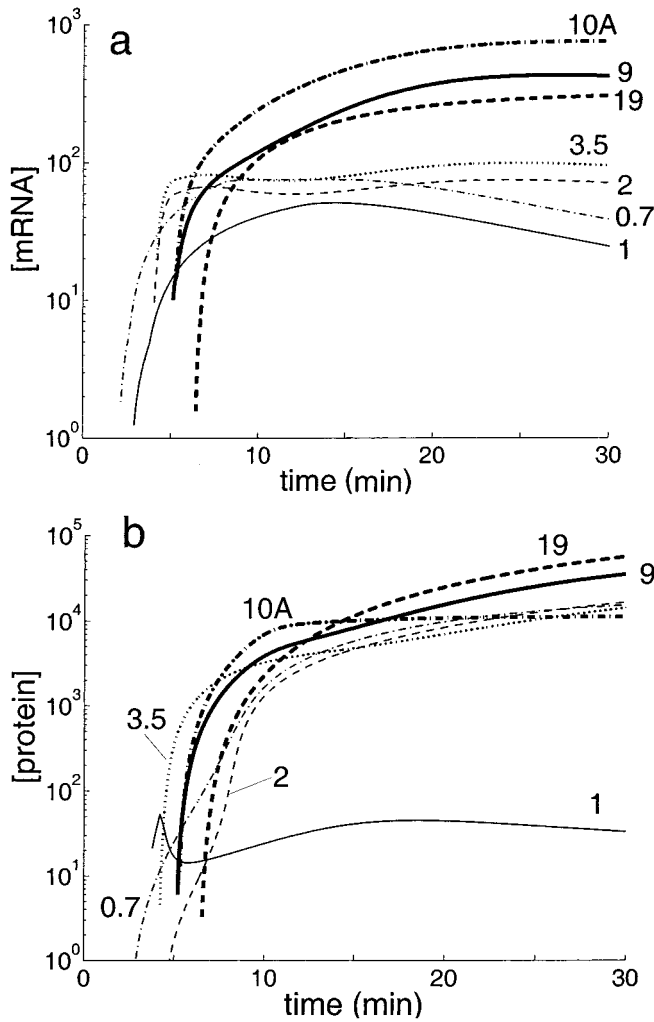


FIG. 2. Time series of (a) mRNAs and (b) free proteins for seven representative T7 genes as labeled.

and gp16, follow similar trends (not shown). gp10A is noteworthy because its maximum of 500 mRNA molecules per cell is the highest achieved by any message, and its protein rate drops to near zero as it approaches this concentration, indicating a strong depletion effect. The other set of nonlinear trajectories including gp0.7, gp1, gp2, and gp3.5 are more complex. The trajectory for gp1, which is the most complex, is shown on expanded axes in Fig. 3b. We know from the simulation that the sudden jump in protein rate from zero to the reference line at low mRNA levels observed for gp1 and other gene products is due to the finite time required by the ribosomes to complete synthesis of the first proteins. The segment of linear increase coinciding with the reference line indicates a growth in mRNA and proteins without any apparent depletion. Then the dramatic drop to

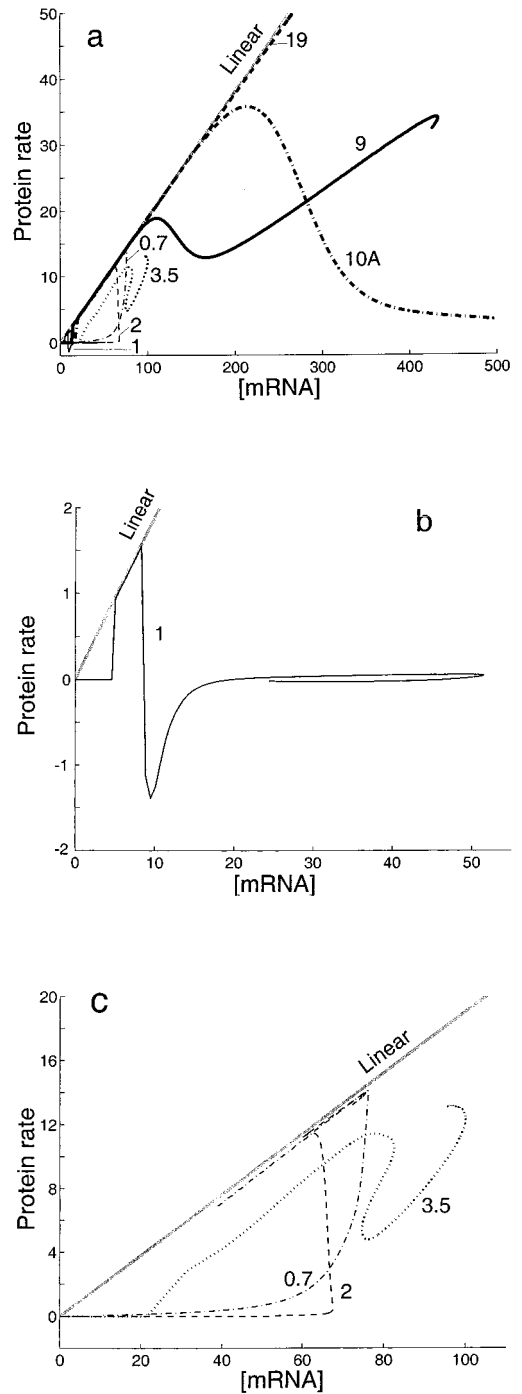


FIG. 3. Protein rate as a function of the corresponding mRNA concentration for (a) seven representative gene products; (b) gp1, expanded axes and (c) gp0.7, gp2, and gp3.5, expanded axes. The protein rate is calculated by taking the first-order finite difference of the corresponding protein concentration, i.e., $v_i \approx \Delta P_i / \Delta t$. The straight line is a plot of v_{Ti} vs R_i , with slope k_T determined directly from the parameters used in the simulation, viz., $k_T = k_E R_d$, where k_E is the ribosomal elongation rate and R_d is the density of ribosomes along the mRNAs (Endy *et al.*, 1997; Endy, 1997).

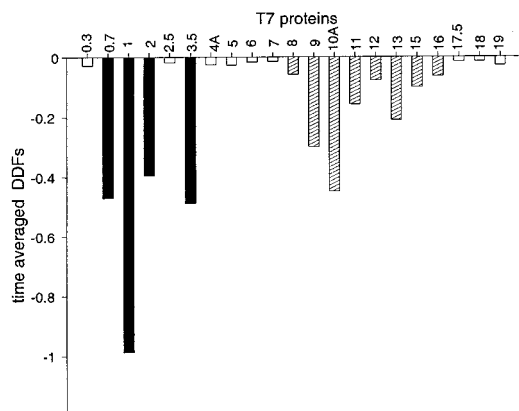


FIG. 4. Time-averaged DDFs for 21 essential T7 proteins. Proteins that exhibit linear (nonshaded) and nonlinear (hatched or black) protein rate versus mRNA trajectories are shown. Gene products that started off linear, but later deviated (hatched) are distinguished from those that started off with large deviations (black).

negative protein rates, even as mRNA is increasing, suggests the appearance of some factor, which could be another component or process, that significantly depletes gp1. After passing through a minimum, the protein rate gradually returns to nearly zero, where it remains, even as the mRNA level goes through a maximum. This pattern suggests that the translation of gp1 is nearly balanced by its depletion during the later stages of development. The trajectories for gp0.7, gp2, and gp3.5 are defined by their extended initial deviation from the reference line, shown in Fig. 3c. Protein rates are near zero well beyond the time required by the ribosomes to complete initial synthesis of each protein, indicating that these proteins are immediately depleted when they otherwise would have begun to appear. The eventual rise and fall of mRNA concentrations along with protein rates trace complex paths that loop back or move along the reference line as gp0.7 and gp2, respectively, or always deviate from the reference line as gp3.5. The different

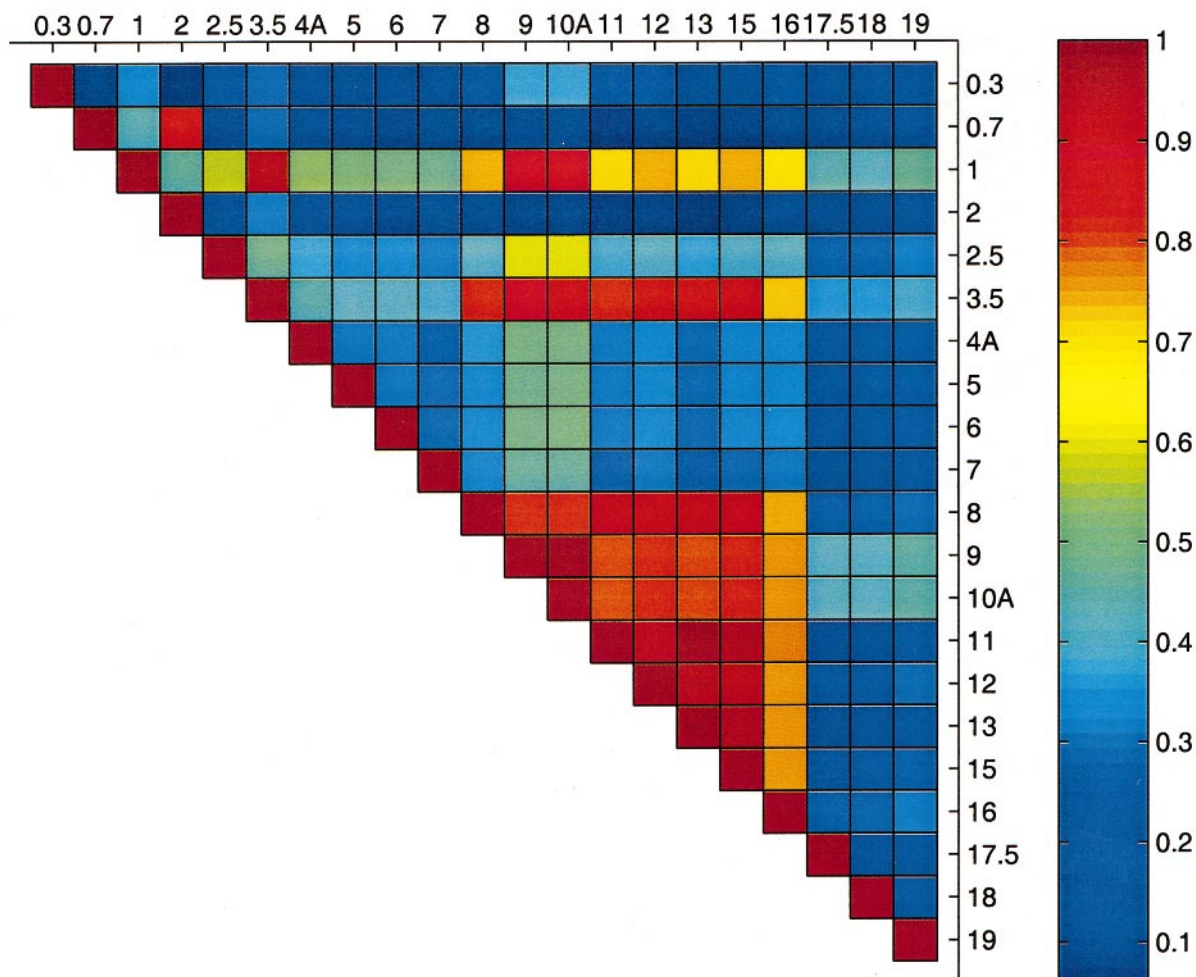


FIG. 5. T7 protein correlation matrix (PCM). Each off-diagonal matrix element represents the correlation coefficient between the DDFs of two proteins, ranging from high (red) to low (blue) correlation. Diagonal elements, which indicate the self-correlation of the proteins, are by definition always equal to 1.

trajectory patterns for 21 essential T7 proteins are summarized in Fig. 4 using time-averaged DDFs.

Using 21 essential T7 proteins, we calculated a protein association matrix for all pairwise interactions. As shown in Fig. 5, high PCCs were found for gp0.7 and gp2, gp1 and gp3.5, and among gene products 8 through 16.

DISCUSSION

We have suggested a correlated deviation algorithm (CDA) for identifying potential patterns of protein function given time series data for mRNAs and proteins. Using data from a genetically structured simulation for the growth of phage T7 we found a variety of expression trajectories, which we classified by their general features, from linear, to nonlinear, to highly nonlinear trajectories that looped back on themselves.

The observed trajectory patterns reflect known functions of the T7 proteins. The complex trajectories of Figs. 3b and 3c describe proteins that all indirectly control their own syntheses, as illustrated by the three negative-feedback loops in Fig. 6. gp0.7 is a protein kinase that is transcribed by EcRNAP, but it also inhibits EcRNAP; increased levels of gp0.7 result in a downregulation of its own transcription. gp2 is an inactivator of EcRNAP, and EcRNAP is required for transcription of T7 RNAP. gp2, however, is predominately transcribed by T7 RNAP, so gp2 ultimately downregulates its own transcription. gp3.5 is a lysozyme that associates with and inhibits T7 RNAP (gp1), but it is also transcribed by T7 RNAP, so it also downregulates its own transcription. Hence, gp0.7, gp2, and gp3.5 share the feature that they all downregulate their own transcription by inhibiting their own RNA polymerases. gp1 is still more involved because it not only downregulates its own transcription through the effect of gp2 on EcRNAP, but also transcribes its own inhibitor in gp3.5.

The protein correlation matrix (PCM) (Fig. 5) is overall consistent with the known functions. The deviations for gp1 and gp3.5 are correlated because they interact with each other, and those of gp0.7 and gp2 are correlated because both proteins are depleted through their interactions with EcRNAP. Moreover, gp8 through gp16, including especially gp9 and gp10A, are highly correlated with each other because they are depleted in a stoichiometric manner during phage particle formation, as shown in Table 1. The gp9 and gp10A trajectories exhibited few if any early deviations (Fig. 3a) because the particle formation process that depletes these proteins occurs late in phage development. With the exception of gp9, gp8 through gp16 are components of the final phage particle. Although gp9 is not present in the final particle, it is required for and consumed by

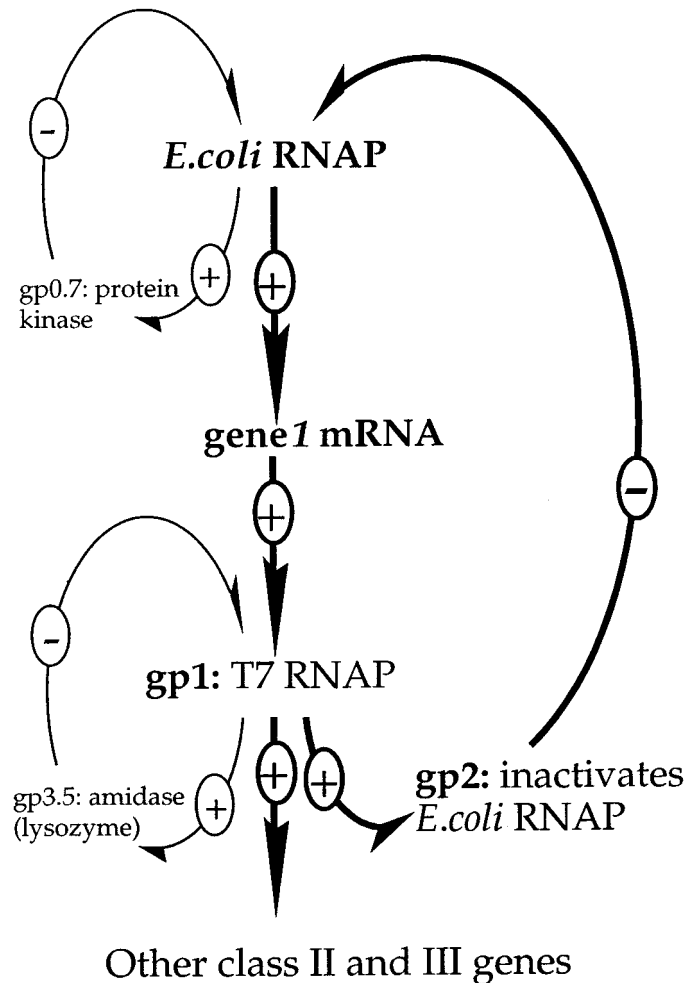


FIG. 6. Negative-feedback loops in the early stages of T7 infection. gp0.7 inhibits EcRNAP via an unknown mechanism, but it is assumed in the simulation that gp0.7 and EcRNAP form a 1:1 complex. gp2 inhibits the EcRNAP by forming a 1:1 complex with the polymerase that prevents transcription. gp3.5 inhibits gp1 by binding the polymerase and increasing the rate of aborted transcript production.

the particle assembly process. Since the matrix provides correlations, we caution against attempting to infer mechanisms from it. Spurious associations for which no interactions were implemented in the simulation, such as those between gp1 and gp3.5 with particle proteins, are evident. Extension of the correlation analysis to allow for time lags may improve discrimination.

The context of the CDA is summarized in Fig. 7. The CDA assumes that the time series of mRNAs and proteins for a given multigene process have sufficient accuracy and resolution for analysis. The protein rate versus mRNA trajectories may reveal patterns that reflect the function of the proteins. The time-averaged DDFs serve as a qualitative measurement of the nonlinear behavior of proteins. The

TABLE 1

Functions and Phage Particle Stoichiometry Data for Selected T7 Proteins (Studier and Dunn, 1983; Steven and Trus, 1986)

Protein	Function or stoichiometry data
gp0.3	Anti restriction protein
gp0.7	Protein kinase
gp1	T7 RNA polymerase
gp2	Inhibitor of <i>E. coli</i> RNAP
gp3.5	Lysozyme, inhibitor of T7 RNAP
gp4A/B	Primase/helicase
gp5	DNA polymerase
gp6	Exonuclease
gp8	Head–tail connection protein (12/phage)
gp9	Head assembly protein (137/phage)
gp10A	Major head protein (415/phage)
gp11	Tail protein (18/phage)
gp12	Tail protein (6/phage)
gp13	Core protein (33/phage)
gp14	Core protein (18/phage)
gp15	Core protein (12/phage)
gp16	Core protein (3/phage)
gp17.5	Lysis protein
gp18	DNA protein
gp19	DNA maturation

PCM further provides a global picture of the potential associations among the proteins of interest. The CDA aims to facilitate the data-mining process of the large volume of output generated from emerging high-throughput experimental techniques, while focusing subsequent experiments on the proteins that demonstrate interesting behaviors.

Reverse engineering approaches have been developed to deduce the underlying wiring of Boolean logic networks, given sufficient state transition (input–output) pairs (Liang *et al.*, 1998; Akutsu *et al.*, 1999). Such approaches may eventually be useful for identifying biological regulatory functions, but to date their application has been limited to the analysis of nonbiological networks.

An important feature of the CDA lies in its use of protein rate versus mRNA trajectories to infer mechanisms. It is an intriguing question to ask whether a “mapping” between particular protein functions and some unique patterns evident in the protein rate versus mRNA trajectories exists. Our analysis with the simulated T7 system suggests this might be the case. We are currently exploring this question by testing the CDA with some simple artificial genetic networks.

The CDA resembles the correlation matrix construction approach proposed by Arkin and Ross (1995) in its use of a correlation matrix, but differences are apparent. Although the PCC is analogous, in both its meaning and form, to the conventional correlation coefficient with zero time lags,

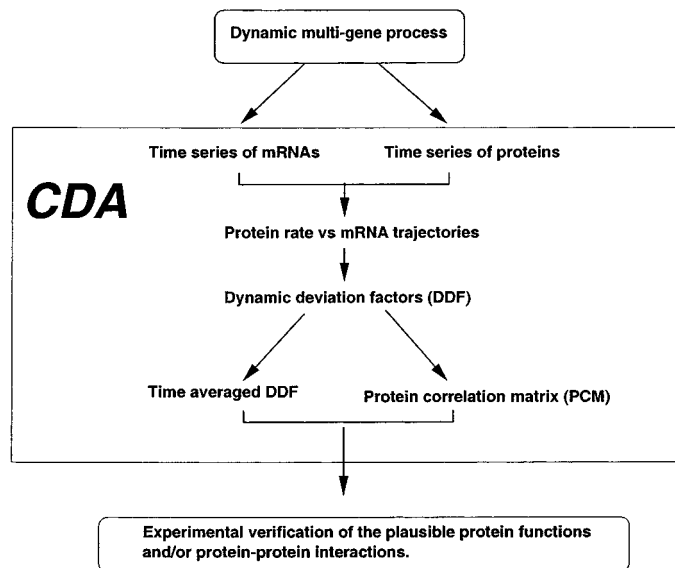


FIG. 7. Context of the correlated deviation algorithm (CDA). By organizing and presenting the data in the forms of protein rate vs mRNA trajectory, DDF, and PCM, the CDA can highlight proteins that demonstrate interesting behavior. These analyses may facilitate processing of output from the emerging high-throughput experimental techniques.

there is significant difference between the two concepts: the conventional correlation coefficient emphasizes the deviation of a variable from its mean when the system is near a steady state or dynamic equilibrium (Arkin and Ross, 1995), but the PCC focuses on the deviation of a protein rate from its “expected” value, and it does not matter whether the system is near a steady state or not.

CONCLUDING REMARKS

Several issues need to be addressed before one can apply the CDA to real data sets. To infer functions from mRNA and protein time series we have used a model system and employed many simplifying assumptions. In general, the magnitude of the translation rate for a given protein species, v_{Ti} , will depend on the distribution of translation resources among messages, the efficiency of the ribosomes, the level of message, the concentration of activated tRNAs, biases in codon usage, and mRNA structure. We simplified our analysis by making translation dependent only on message levels, but assumed uniform translation rates for all messages and no limitations on ribosomes. We have explored the effects of ribosome limitations and obtained similar results, while assuming that the limited ribosomes were uniformly distributed across all messages (not shown). We have neglected protein degradation, modification, or transport out of the cytoplasm, leaving the formation of protein–protein interactions as the lone mode for depleting a protein. These assumptions are reasonable for the phage T7

model system, but they will need to be addressed for more complex systems. Finally, we have neglected stochastic effects due to the indeterminate behaviors of small numbers of molecules, effects that may be especially magnified when transcripts and proteins are just beginning to appear (Gillespie, 1977; McAdams and Arkin, 1997).

This work has focused on concepts rather than practice. We have saved for later questions regarding the data quality or frequency of data sampling needed to discriminate among different classes of trajectories. Moreover, although DNA-array technologies are beginning to provide global profiles of mRNA over time (Spellman *et al.*, 1998; DeRisi *et al.*, 1997), such profiles for free protein present both a significant technical challenge and an opportunity. Global protein profiles obtained by two-dimensional gel electrophoresis are typically carried out under conditions that destroy noncovalent protein–protein interactions (Fichmann and Westermeier, 1999), which also destroys information about the free protein levels that we have used in our analysis. Current technologies to detect and analyze protein–protein interactions focus explicitly on forming and studying the protein–protein complex (Phizicky and Fields, 1995). Our analysis suggests an opportunity to indirectly infer protein–protein interactions under physiological conditions if free proteins can be isolated from the cell, separated, and measured without disrupting existing complexes.

ACKNOWLEDGMENTS

We thank the Office of Naval Research (Grant N00014-98-1-0226) and the National Science Foundation Presidential Early Career Award to J.Y. (Grant BES-9896067) for supporting this work.

REFERENCES

- Akutsu, T., Miyano, S., and Kuhara, S. (1999). Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Pac. Symp. Biocomput.*, 17–28.
- Arkin, A., and Ross, J. (1995). Statistical construction of a chemical reaction mechanism from measured time-series. *J. Phys. Chem.* **99**, 970–979.
- Arkin, A., Shen, P., and Ross, J. (1997). A test case of correlation metric construction of a reaction pathway from measurements. *Science* **277**, 1275–1279.
- Dayton, C. J., Prosen, D. E., Parker, K. L., and Cech, C. L. (1984). Kinetic measurements of *Escherichia coli* RNA polymerase association with bacteriophage T7 early promoters. *J. Biol. Chem.* **259**, 1616–1621.
- DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.
- Dunn, J. J., and Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**, 477–535.
- Eckerskorn, C., Strupat, K., Karas, M., Hillenkamp, F., and Lottspeich, F. (1992). Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization. *Electrophoresis* **13**, 664–665.
- Endy, D. (1997). “Development and Application of a Genetically Structured Simulation for Bacteriophage T7,” Ph.D. thesis, Dartmouth College.
- Endy, D., Kong, D., and Yin, J. (1997). Intracellular Kinetics of a growing virus: A genetically structured simulation for bacteriophage T7. *Biotech. Bioeng.* **55**, 375–389.
- Fichmann, J., and Westermeier, R. (1999). 2-D protein gel electrophoresis. An overview. *Methods Mol. Biol.* **112**, 1–7.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361.
- Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* **90**, 5011–5015.
- Kusakabe, T., Baradaran, K., Lee, J., and Richardson, C. C. (1998). Roles of the helicase and primase domain of the gene 4 protein of bacteriophage T7 in accessing the primase recognition site. *EMBO J.* **17**, 1542–1552.
- Liang, S., Fuhrman, S., and Somogyi, R. (1998). Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.*, 18–29.
- McAdams, H. H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci.* **94**, 814–819.
- O’Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007–4021.
- Phizicky, E. M., and Fields, S. (1995). Protein–protein interactions: Methods for detection and analysis. *Microbiol. Rev.* **59**, 94–123.
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Po, P. O. B., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297.
- Steven, A. C., and Trus, B. L. (1986). The structure of bacteriophage T7, in “Electron Microscopy of Proteins” (J. R. Harris and R. W. Horne, Eds.), Vol. 5, pp. 1–35, Academic Press, London.
- Studier, F. W., and Dunn, J. J. (1983). Organization and expression of bacteriophage T7 DNA. *CSH Quant. Biol.* **47**, 999–1007.
- Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* **270**, 484–487.