

Experimental Design in Genetic Regulatory Network Identification: Results from *In Silico* Studies

Daniel E. Zak^{a,b}, Francis J. Doyle III^{c1}, Ronald K. Pearson^d and James S. Schwaber^b

^aDepartment of Chemical Engineering, University of Delaware, Newark, DE 19716

^bDepartment of Pathology, Cell Biology and Anatomy, Thomas Jefferson University
Philadelphia, PA 19107

^cDepartment of Chemical Engineering, University of California, Santa Barbara, CA 93106

^dTampere University of Technology, Tampere, FIN-33101, Finland

Key words: genetic regulatory networks, experimental design,
modulating functions, system identification

Prepared for Presentation at the 2002 Annual Meeting, Indianapolis, IN, Nov. 3–8
Copyright ©2002, D.E. Zak, University of Delaware & Thomas Jefferson University
F.J. Doyle III, University of California at Santa Barbara
R.K. Pearson, Tampere University of Technology
and James S. Schwaber, Thomas Jefferson University

November 2002

Unpublished

AICHE shall not be responsible for statements or opinions contained in papers or in its publications.

¹Author to whom correspondence should be addressed: doyle@engineering.ucsb.edu

Abstract

As stated by Leroy Hood [11], the difference between man and monkey is gene regulation. One of the great prospects in the post-genomic era is that the genetic regulatory networks that govern such varied processes as development, differentiation, and neuronal adaptation will be uncovered through the analysis of gene expression and related perturbation time courses. In previous work we developed a 10-gene genetic regulatory network simulator for the purpose of determining the suitability of gene expression data for network identification and to benchmark network identification approaches. Using this simulator, we have demonstrated the value of additional datasets that constrain possible gene-gene interactions in addition to gene expression data for the purpose of identifying genetic regulatory networks [22].

In the present study we use our genetic regulatory network simulator to explore the impact of several experimental design variables on the ability of identification methods to identify genetic regulatory networks from gene expression data. Specifically, we consider how excitation in the perturbation sequence (e.g., pulse train versus steps) can influence the results of the identification methods. We also consider the impact of sampling rate and number of samples, which are especially important due to the varied time scales present in genetic regulatory networks and the difficulty in obtaining large numbers of gene expression measurements. Finally, treating individual trajectories of the stochastic simulator as individual cells, we explore the impact of cell sample size on the success of genetic regulatory network identification methods. The network identification method we use is a continuous time power-law model with delay combined with a modulating function approach for parameter estimation.

1 Introduction

The discovery of the regulatory networks governing fundamental cellular processes, such as differentiation, neuronal adaptation, and drug responsiveness, is one of the great promises of microarray and other functional genomic technologies. Given this great potential, the problem of identifying genetic regulatory networks from microarray and similar data has been considered by many authors (see [16, 4, 20, 22, 9, 21], for example).

It is often the case that the investigators that generate microarray data are not the investigators involved in identifying regulatory networks from the data. This situation may not be optimal for the difficult task of network identification, and it may be expected that the groups that are most successful in identifying regulatory networks are those that can directly influence the type of data that is collected. In this case, the design of the microarray experiments for network identification will be important.

In the present work, an *in silico* genetic regulatory network model is used to explore aspects of experimental design that are relevant to genetic regulatory network identification from microarray data. The model is an updated version of a 10-gene, 55-state model system presented previously [22] and is a reasonable starting point for simulating the mechanistic complexity and time scale variations inherent to genetic regulatory networks.

The present paper is structured as follows: the model used for *in silico* experimentation is first described, followed by aspects of experimental design that are relevant to genetic regulatory network identification. The modulating function-based approach to network identification is presented, and followed by the results from *in silico* experimentation.

2 Model for *in silico* experimentation

The model used for *in silico* experimentation has been presented previously [22] and is shown schematically in Figure 1-a. It is a 10-gene network with a receptor that responds to ligand input. It is described

mathematically by 55 coupled nonlinear ordinary differential equations. The transcription of each gene is governed by a transcriptional regulatory module, shown in Equation 1.

$$\begin{aligned}
 \dot{[Pij]} &= -k_{Pij}[Pij][j_2] + k_{uPij}[j_2Pij] \\
 \dot{[j_2Pij]} &= k_{Pij}[Pij][j_2] - k_{uPij}[j_2Pij] \\
 \dot{[i]} &= k_{Ti}[Mi] - 2k_{i2}[i]^2 + 2k_{ui2}[i_2] - k_{di}[i] \\
 \dot{[i_2]} &= k_{i2}[i]^2 - k_{ui2}[i_2] - k_{di2}[i_2] - \Sigma PB + \Sigma PU \\
 \dot{[Mi]} &= k_{RPik}[Pijk] + k_{RjPik}[j_2Pijk] + k_{RkPik}[k_2Pijk] + k_{RjkPik}[k_2j_2Pijk] - k_{dMi}[Mi]
 \end{aligned} \tag{1}$$

where :

k_i = rate constant for reaction i ; $[M_i]$ = conc. of transcript i , $[i]$ = conc. of protein i ;

$[i_2]$ = conc. of active TF i_2 ; $[Pij]$ & $[Pijk]$ = unbound promoter conc.;

$[j_2Pij]$, $[j_2Pijk]$, $[k_2Pijk]$, & $[j_2k_2Pijk]$ = bound promoter conc.

The overall architecture of the network is shown in Figure 1-a. This nonlinear dynamical system may be abstracted into a qualitative matrix that contains the signs of direct gene–gene and ligand (Q)–gene interactions, with the *inputs* influencing the transcription rates of the *outputs* (Figure 1–b). This abstraction is only a partial representation of the complete system, but in most genetic regulatory network identification studies this is what is sought.

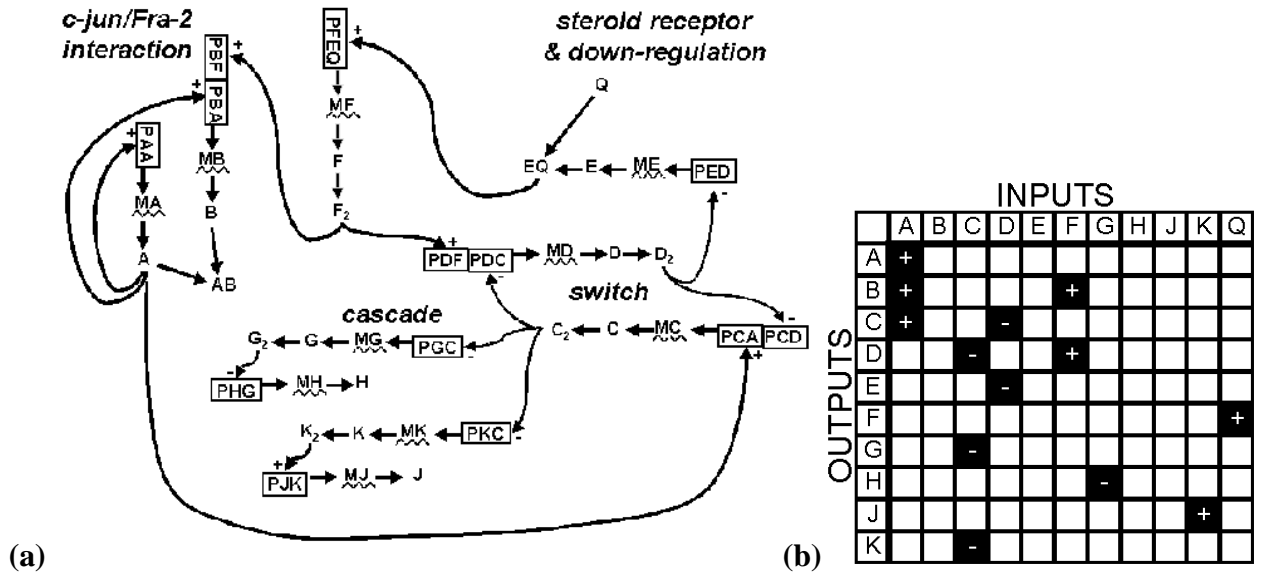


Figure 1: (a) Overall model architecture. (b) Abstracted gene–gene interaction matrix. + = transcriptional activation, – = transcriptional repression

The model has been modified since its original presentation to include literature–based time scales representative of mammalian gene expression. A typical deterministic expression profile in response to a prolonged pulse of ligand is shown in Figure 2–a, where the varied time scales on which the genes respond to the perturbation are apparent. Note that the type of ligand input (step, pulse, etc.) used to excite the system is an aspect of experimental design that may be readily explored with this model. Integrations were performed using the stiff solver *ode15s* in MATLAB (The Mathworks, Inc.).

A novel hybrid stochastic/deterministic simulation approach that stochastically integrates transcripts (M_i) and promoters (P_{ij}) using Gillespie’s Direct Method [7] while integrating the proteins and transcription factors deterministically has been used to generate stochastic data from this model. Simulating aspects of

gene expression in a stochastic framework is important because there is accumulating experimental evidence that transcription is a stochastic process [13, 5]. Stochastic simulations allow another aspect of experimental design to be explored: N_{cells} , the number of cells collected per sample. This is important because single-cell expression profiling is a reality [10]. It should be noted that a single stochastic time course does not represent a single-cell gene expression time course, since current methods for measuring gene expression are destructive to cells. A more correct approach to simulating the effect of stochasticity in gene expression on microarray time courses is to generate at least $N_{samp} \times N_{cells}$ stochastic time courses, where N_{samp} is the number of time points sampled. For each time point, N_{cells} stochastic time courses or *cells* are selected randomly from the total set. The expression level at each time point is then the sum of the expression levels of the randomly selected *cells* at that time point. This procedure is illustrated in Figure 2–b, where $N_{cells} = 1$ and $N_{samp} = 5$. Note that the stochasticity of the gene expression may give a time course that is qualitatively different from any of those in the original cells.

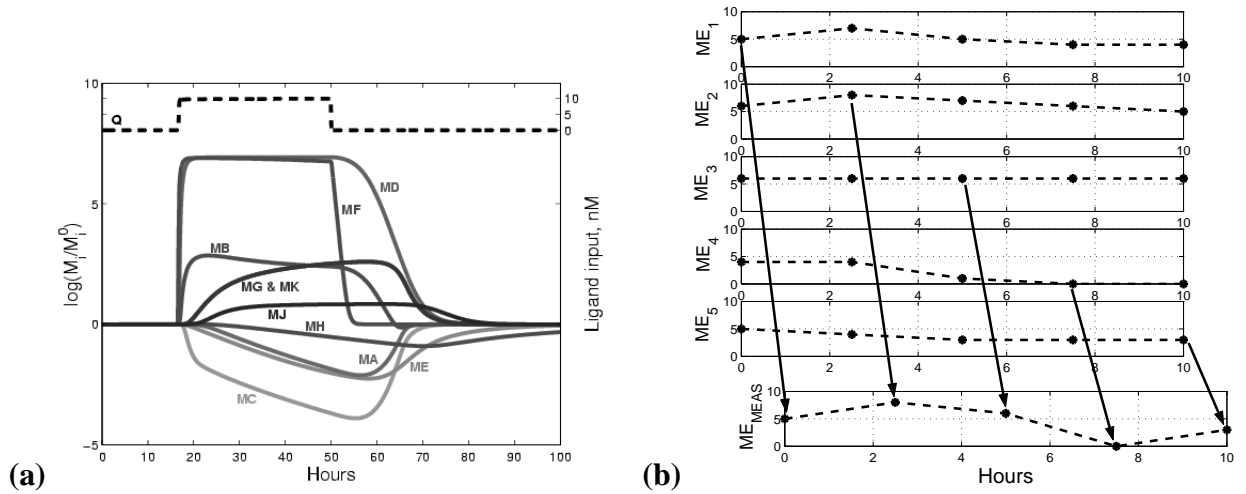


Figure 2: (a) Log(expression ratios) versus time for all genes (colored lines) and ligand (dashed line) from a deterministic simulation, (b) example of how multiple stochastic simulations are used to create a microarray time course that includes intrinsic stochasticity in gene expression, gene E, $N_{cells} = 1$, $N_{samp} = 5$. The composite expression profile (ME_{MEAS}) is qualitatively different from the profiles of the original cells

3 Experimental design

Following Walter and Pronzato (1990), experimental design has qualitative and quantitative aspects [18].

Qualitative experimental design: Qualitative aspects of experimental design are concerned with selecting the correct measurements to make the parameters of interest identifiable. What information has been collected prior to the microarray experiment may be considered a qualitative aspect of experimental design for network identification. Two cases of prior knowledge are considered presently:

- *Transcript degradation rate constants:* These govern the time scales on which genes respond to changes in transcription [8]. They are an increasingly available type of data [19, 6], and the importance of regulation of gene expression through transcript turnover is increasingly recognized [6].
- *Localization information:* Localization information specifies *a priori* which transcription factors bind to which genes and may be obtained without gene expression profiles. It is also becoming an increasingly important data source [17] and it has been recognized that network identification is greatly improved when this information is used in conjunction with microarray data [22, 9].

Quantitative experimental design: Quantitative aspects of experimental design concern optimizing the information extracted from the experimental data by adjusting key experimental design variables. The experimental design variables considered presently are:

- *Richness of perturbation:* For the present model, this is the dynamic structure of the ligand input. Single step, a prolonged pulse (2 step), single pulse, and pulse + step ligand inputs are considered presently. The default input is the single pulse (time 0–28 hrs in Figure 2–a).
- *Sampling range:* For the pulse and step–chase experiments considered presently, this involves selecting t_{fin} , the final time at which samples are taken. The default value of t_{fin} is 28hrs.
- *Sampling protocol:* For fixed interval sampling, samples must be taken at a frequency at least twice the highest frequency in the system to avoid aliasing [2]. In nonlinear systems, this frequency may change over time and is difficult to define. Kalogerakis and Luus [12] suggest a log–linear sampling protocol to capture multiple time scales. Both the fixed interval and log–linear sampling protocols are considered in the present work.
- N_{samp} : This is the number of samples taken. The nominal value of N_{samp} is 10 samples.
- N_{cells} : This is the number of cells that are sampled per time point (see previous section). The nominal value of N_{cells} is 20 cells.

4 Modulating function approach to network identification

A modulating function approach is employed for the identification of the gene–gene and ligand–gene interaction parameters. The advantages of the modulating function approach are that it removes any need for approximating derivatives from data, there is no need to estimate initial conditions, and there is an inherent data filtering characteristic, allowing models to be fit using noisy data [3]. The drawback is that the approach requires the models to be linear in the parameters. In the present work, the regression for each gene is carried out individually. This leads to N MISO (multiple input, single output) problems, where N is the number of genes in the system.

In a similar manner to D’haeseleer *et al.* [4], the gene–gene and ligand–gene interaction parameters for each gene are desired. These may be represented by a_{ij} and b_{ik} for gene i in:

$$\dot{x}_i = \sum_{j=1}^J a_{ij} f_1\{x_j(t - \tau_1)\} + \sum_{k=1}^K b_{ik} f_2\{q_k(t - \tau_2)\} + c_i - k_{Di} x_i \quad (2)$$

where $\mathbf{x}(t)$ are expression levels for all genes over time obtained from the microarray data, $\mathbf{q}(t)$ are levels for all ligands over time, and $f_1\{\cdot\}$ and $f_2\{\cdot\}$ are known functions that approximate how transcript and ligand levels influence the transcription of each gene. Treating the nonlinear terms involving f_1 and f_2 as inputs [3] and changing indices so that $L = J + K$ gives:

$$\dot{x}_i = \sum_{l=1}^L g_{il} u_l(t - \tau_l) + c_i - k_{Di} x_i \quad (3)$$

The parameters in Equation 3, $[g_{i1}, g_{i2}, \dots, g_{iL}, c_i, -k_{Di}]^T$, may be identified from a time series of $\mathbf{x}(t)$ and $\mathbf{u}(t)$ by linear regression of:

$$- \begin{bmatrix} A_{i_1} \\ A_{i_2} \\ A_{i_3} \\ \vdots \\ A_{i_{M-1}} \\ A_{i_M} \end{bmatrix} = \begin{bmatrix} B_{i1_1} & B_{i2_1} & \dots & B_{iL_1} & C_{i_1} & D_{i_1} \\ B_{i1_2} & B_{i2_2} & \dots & B_{iL_2} & C_{i_2} & D_{i_2} \\ B_{i1_3} & B_{i2_3} & \dots & B_{iL_3} & C_{i_3} & D_{i_3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{i1_{M-1}} & B_{i2_{M-1}} & \dots & B_{iL_{M-1}} & C_{i_{M-1}} & D_{i_{M-1}} \\ B_{i1_M} & B_{i2_M} & \dots & B_{iL_M} & C_{i_M} & D_{i_M} \end{bmatrix} \begin{bmatrix} g_{i1} \\ g_{i2} \\ \vdots \\ g_{iL} \\ c_i \\ -k_{Di} \end{bmatrix} \quad (4)$$

where:

$$\begin{aligned}
A_{i_z} &\triangleq \int_0^{t_{fin}} \dot{\phi}_z x_i dt \\
B_{i_z} &\triangleq \int_0^{t_{fin}} \phi_z(t + \tau_l) u_l(t) dt + \tau_l \{ \phi_z(\tau_l) u_l(0) - \phi_z(t_{fin} + \tau_l) u_l(t_{fin}) \} / 2 \\
C_{i_z} &\triangleq \int_0^{t_{fin}} \phi_z dt \\
D_{i_z} &\triangleq \int_0^{t_{fin}} \phi_z x_i dt
\end{aligned} \tag{5}$$

and where the ϕ_z are linearly independent differentiable modulating functions with the property: $\phi_z(0) = \phi_z(T) = 0$. Following Co and Ydstie [3] and Pearson and Lee [14], trigonometric functions are used in the present study. Two specific classes are used: $\phi_{z_1} = \sin(n\omega t/2)$ and $\phi_{z_2} = \cos(n\omega t) - \cos((n-1)\omega t)$, where $\omega \triangleq 2\pi/t_{fin}$. A total of 12 modulating functions were used for all cases ($n = 1 : 8$ for ϕ_{z_1} and $n = 1 : 4$ for ϕ_{z_2}). Note that to arrive at Equation 5 from the properties of ϕ_z and Equation 3, a one-step trapezoidal approximation was used for some integrals involving the delayed inputs.

To allow for non-uniform sampling without significant integration error, the integrations in Equation 5 were performed by first linearly interpolating between data points, analytically integrating the product of the modulating functions and the interpolated lines, and then summing over all time points:

$$\int_0^{t_{fin}} \phi(t)x(t)dt \approx \sum_{j=1}^{N_{samp}-1} \frac{1}{t_{j+1}-t_j} \int_{t_j}^{t_{j+1}} \phi(t)\{x_{j+1}(t-t_j) + x_j(t_{j+1}-t)\}dt \tag{6}$$

In the present study, a simplified version of Equation 2 is used for network identification:

$$\dot{x}_i = \sum_{j=1}^2 a_{ij}x_j(t-\tau)^{0.5} + b_{i1}q_1(t-\tau)^{0.5} + c_i - k_{D_i}x_i \tag{7}$$

where three cases are considered: (1) each gene is influenced by one gene (ligand), (2) each gene is influenced by at most two genes (ligand), or (3) the gene-gene interactions are known *a priori* (localization information). Fractional exponents are used to approximate saturating kinetics, in a similar manner as Savageau's approximation of complex mechanisms with power-law kinetics [15]. It has been assumed that the gene-gene and ligand-gene delays are equal for all genes, taken to be 0.5 hrs. Allowing gene-dependent time delays and identifying the delays from data [1] will be pursued in future work.

5 Validation of network identification: *correctness and fitting*

To validate the network identification results it is convenient to introduce a measure (Δ) that quantifies the success of the identification. For gene $_i$:

$$\Delta_i \triangleq \sum_j |sign(sign(g_{ij}^{corr}) - sign(g_{ij}^{pred}))| \tag{8}$$

where g_{ij}^{corr} are the signs of the "correct" interactions (ex: Figure 1-b), in the sense that they are signs of the gene-gene and ligand-gene transcriptional interactions that are known to exist in the model. g_{ij}^{pred} are the best fit parameters obtained using the modulating function approach described above. For the entire network, $\Delta_{overall} \triangleq \sum_i \Delta_i$. An identification method that gives a small Δ successfully identifies the transcriptional interactions in the network. In other words, Δ is inversely proportional to the "correctness" of the network identification.

Δ can be compared against a measure of the *goodness of fit*. Goodness of fit can be quantified by Γ , the sum of the sum of squared errors (SSE) of the best fits for all genes: $\Gamma \triangleq \sum_i \gamma_i$, where γ_i is the SSE of the best fit input for gene i . In other words, Γ is inversely proportional to the *goodness of fit* of the identification model.

6 *In silico* experiments

In the following, results from *in silico* experiments are presented. Several aspects of experimental design relevant to network identification were explored, including the impact of prior knowledge (transcript degradation rate constants and localization information), the time range over which samples were taken (t_{fin}), the number of samples (N_{samp}), the number of cells collected per sample (N_{cells}), and the perturbation richness (ligand single step, two steps, single pulse, and pulse + step). The nominal data set used was the step response shown in Figure 2–a, 0–28hrs. Both deterministic (without noise) and stochastic (20 cells) simulation data was used. The sampling protocol (fixed interval or log-linear [12]) was found to have a negligible effect on the network identification results (not shown).

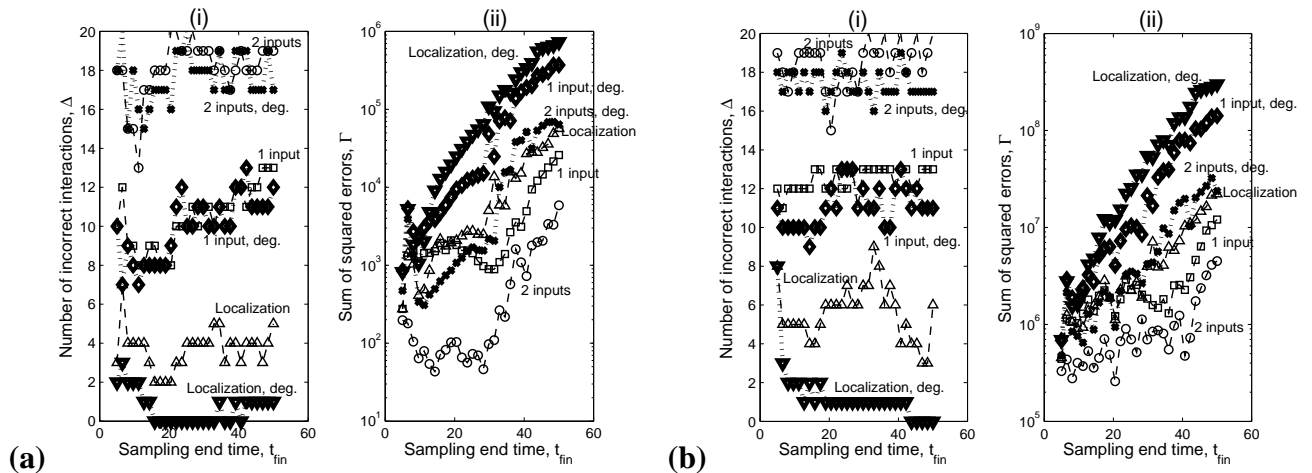


Figure 3: Effect of sampling range (t_{fin}) on correctness (Δ) and fit (Γ) of identification. **(a) Deterministic data**, (i) The optimal range of t_{fin} for correctness (min Δ) for cases without localization information is small and narrow, while it is wider for cases with localization information. (ii) The optimal range of t_{fin} for fitting (min Γ) roughly corresponds to the range for optimal correctness. **(b) Stochastic data**, (i) the correctness is much less sensitive to t_{fin} than for deterministic data. (ii) the fit decreases monotonically with increasing t_{fin} . Note that for all cases, prior localization information drastically improved the identification. Prior transcript degradation information improved the identification only for cases with prior localization information.

7 Results and conclusions

From the results in Figures 3, 4, 5, and 6, several conclusions can be drawn:

- For the network identification approach considered, the sampling protocol was not an important aspect of experimental design.
- The optimal time window was found to be small for the deterministic simulations, possibly indicating a region where the power-law approximation was most valid.
- Increasing samples or cells improves the identification only when few samples are taken or few cells/sample are collected.
- Ligand inputs that are more dynamically rich do not necessarily lead to a close correspondence between the signs of interactions in the best fit model and the signs of transcriptional interactions in the underlying system. In some cases the less rich inputs (single pulse and single step) outperformed the richer inputs.
- For all cases, when localization information was available, the signs of interactions in the best fit models corresponded closely with the “correct” signs of the interactions. For this reason, prior localization information combined with gene expression profiles allows robust identification of the

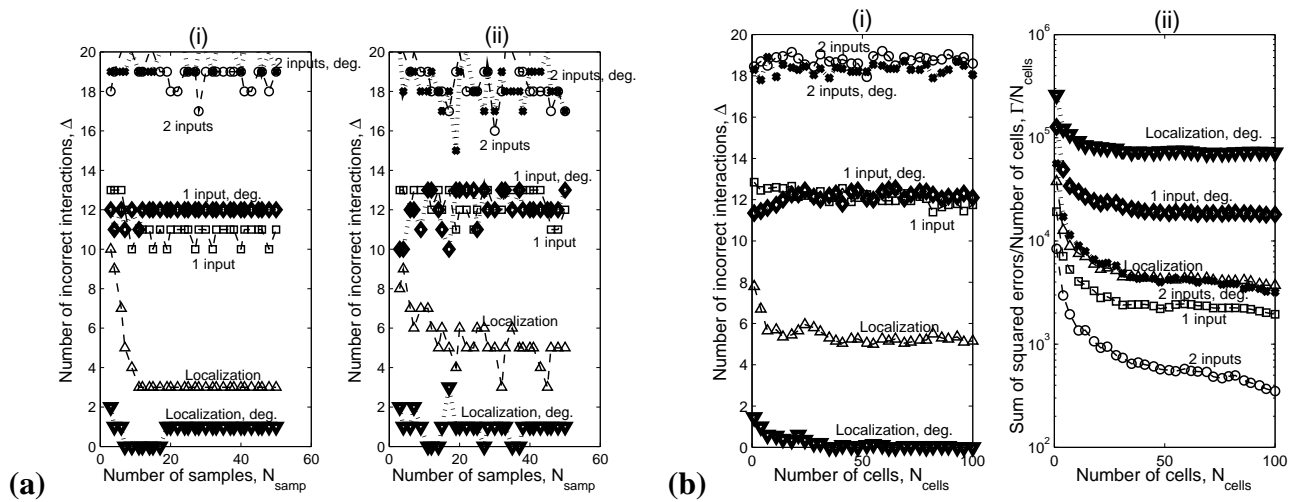


Figure 4: (a) **Correctness (Δ) versus N_{samp}** , for deterministic (i) and stochastic (ii) data. For both cases, the success of the identification is not very sensitive to N_{samp} , while N_{samp} is important at low values for the cases with localization. (b) **Correctness (Δ) and fit (Γ/N_{cells}) versus N_{cells}** , average of 20 runs. (i) Success of identification is dependent on N_{cells} only for low values. (ii) As more cells are collected per sample, the stochastic fluctuations become less significant, and the sum of squared errors/cell decreases. Note that the decrease is not monotonic, as the improvement in fit levels off around 20 cells/sample.

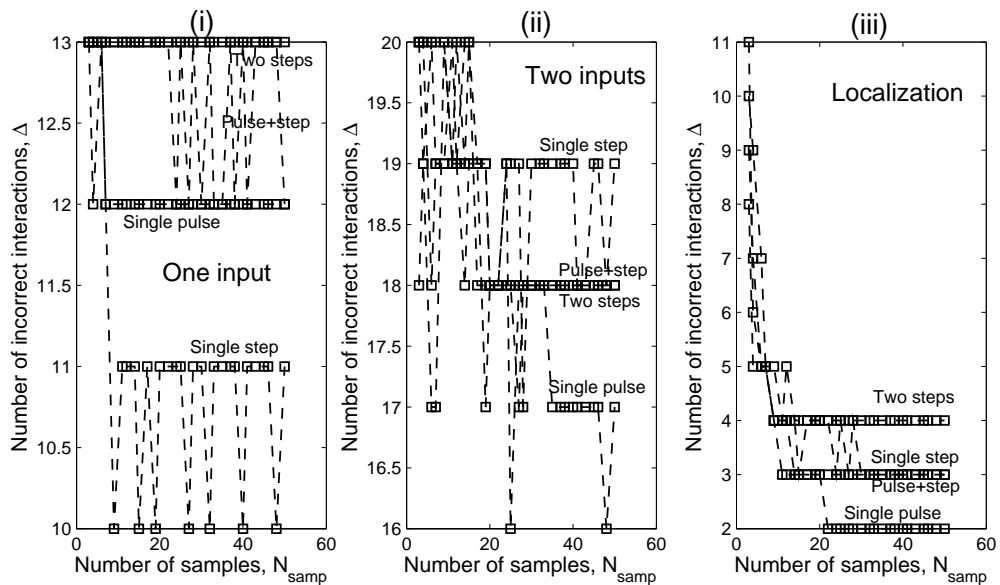


Figure 5: Success of network identification versus N_{samp} for multiple ligand inputs, transcript degradation rates not known, deterministic data. (i) Single input/gene, (ii) 2 inputs/gene, (iii) localization information. The single step gives the best network identification (min Δ) for (i), while the single pulse is best for (ii) and (iii).

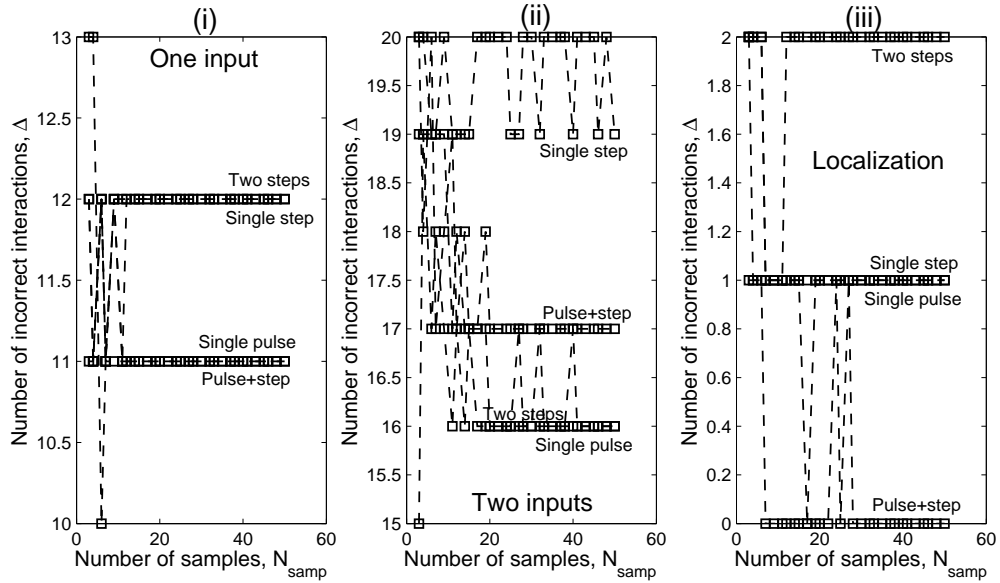


Figure 6: Success of network identification versus N_{samp} for multiple ligand inputs, transcript degradation rates known, deterministic data. (i) Single input/gene, (ii) 2 inputs/gene, (iii) localization information. The single-pulse and pulse+step give the best network identification (min Δ) for A, single-pulse and two-steps are best for B, and pulse+step is best for C.

network. It must be noted, however, that the cases with localization information gave significantly worse fits (higher SSE) than cases without.

- Only for cases with prior localization information did prior knowledge of the degradation rate constants significantly improve the identification result. For most cases, the subsequent improvement was not as significant as the original improvement from including localization information.

In the present work, experimental design for genetic regulatory network identification has been explored using the ten-gene biological network simulator. Similar to other studies [22, 9], it was found that localization information improves genetic regulatory network identification from microarray data. Including transcript degradation rate constants improved the identification even further. These results are important because, in the present time frame, only a few microarray datasets with significant noise are likely to be collected in any particular experiment. Identifying which additional datasets that may combined with the limited microarray data to make network identification tractable is an important step.

There were also counter-intuitive results concerning the relationship between the best fit (smallest sum of squared errors) and most correct (smallest number of incorrect signs) networks. The cases with prior localization information always had significantly worse fits than cases without, even though they most correctly identified the underlying network. This is important because cases with localization information automatically have the correct structure and only the parameters need to be fit. It is expected that networks with correct structures would fit the data better than cases with incorrect structures; it is surprising that they did not. This result may be due to the significant differences between the system that generated the data and the systems that were identified from the data. It may be that the simple power-law with delay models are not flexible enough to accurately capture the dynamics of the true system. Another possibility is that the parameters that determine the network architecture (transcription factor-promoter interactions and transcription rates with bound/unbound promoters) may not be identifiable from microarray data alone. Using a small genetic regulatory network model, it has been observed that this may be the case (*unpublished results*). This may also explain why increasing the data richness did not necessarily improve

the identification results, since it may be *a priori* impossible to determine a unique network structure from microarray data alone. Rigorous identifiability analyses are necessary to verify this hypothesis. If it is disproved, finding models that are simple enough that they may be identified with confidence from limited data while still being able to capture the complexities underlying real genetic networks will be a challenge that must be met for successful genetic regulatory network identification from microarray data alone.

Acknowledgments: We thank NIH/NHLBI (R01 HL54194-05), NIH/NIAAA IRPG (R01 AA-13204-01), NIH/NIGMS (P20 H64459) and DARPA BioCOMP (F30602-01-2-0578) for funding. DEZ also acknowledges NIH training grant NIAAA5T32AA07463-15 for funding.

References

- [1] A. Balestrino, A. Landi, and L. Sani. Identification of Hammerstein systems with input/output time delay via modulating functions. In *Proc. IFAC. LTDS2000*, pages 168–172, September 2000.
- [2] L. Balmer. *Signals and Systems – an Introduction*. Prentice Hall, 1991.
- [3] T. B. Co and B. E. Ydstie. System identification using modulating functions and fast fourier transforms. *Comput. Chem. Eng.*, 14(10):1051–1066, 1990.
- [4] P. D’Haeseleer, X. Wen, S. Fuhrman, and R. Somogyi. Linear modeling of mRNA expression levels during CNS development and injury. In *Pacific Symposium on Biocomputing 4*, pages 41–52, 1999.
- [5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.
- [6] J. Fan, X. Yang, W. Wang, W. H. Wood, K. G. Becker, and M. Gorospe. Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proc. Natl. Acad. Sci. USA*, 99(16):10611–10616, 2002.
- [7] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.
- [8] J. L. Hargrove. Microcomputer-assisted kinetic modeling of mammalian gene expression. *FASEB J*, 7:1163–1170, September 1993.
- [9] A. J. Hartemink, D. K. Gifford, T. S. Jaakola, and R. A. Young. Combining location and expression data for principled discovery of genetic regulatory network models. In *Proc. Pac Symp Biocomput*, 7:437–49, 2002.
- [10] S. E. Hemby, S. D. Ginsberg, B. Brunk, S. E. Arnold, J. Q. Trojanowski, and J. H. Eberwine. Gene expression profile for schizophrenia. *Arch. Gen. Psychiat.*, 59(7):631–40, 2002.
- [11] L. Hood. In *Bio2001, San Diego, CA.*, June 2001.
- [12] N. Kalogerakis and R. Luus. Sequential experimental design of dynamic systems through use of the information index. *Can. J. Chem. Eng.*, 62:730–737, 1984.
- [13] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nat. Genet.*, 31(1):69–73, 2002.
- [14] A. E. Pearson and F. C. Lee. On the identification of polynomial input–output differential systems. *IEEE T. Automat. Contr.*, AC-30(8):778–782, 1985.
- [15] M. A. Savageau. Biochemical systems theory: operational differences among variant representations and their significance. *J. Theor. Biol.*, 151(4):509–30, 1991.
- [16] D. H. Sharp and J. Reinitz. Prediction of mutant expression patterns using gene circuits. *Biosystems*, 47(1–2):79–90, 1998.
- [17] I. Simon, J. Barnett, N. Hannett, C. T. Harbison, N. J. Rinaldi, T. L. Volkert, J. J. Wyrick, J. Zeitlinger, D. K. Gifford, T. S. Jaakkola, and R. A. Young. Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, 106:697–708, 2001.
- [18] E. Walter and L. Pronzato. Qualitative and quantitative experiment design for phenomenological models – a survey. *Automatica*, 26(2):195–213, 1990.
- [19] Y. Wang, C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA*, 99(9):5860–5865, 2002.
- [20] L. F. Wessels, E. P. V. Someren, and M. J. Reinders. A comparison of genetic network models. In *Pacific Symposium on Biocomputing 6*, pages 508–519, 2001.
- [21] M. K. Yeung, J. Tegner, and J. J. Collins. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl. Acad. Sci. USA*, 99(9):6163–6168, 2002.
- [22] D. E. Zak, F. J. Doyle III, G. E. Gonye, and J. S. Schwaber. Simulation studies for the identification of genetic networks from cDNA array and regulatory activity data. In *Proc. 2nd Intl. Conf. Systems Biology*, pages 231–238, November 2001.