

# Local identifiability: when can genetic networks be identified from microarray data?

Daniel E. Zak  
Department of Chemical  
Engineering, University of  
Delaware  
& DBI, Thomas Jefferson  
University  
Philadelphia, PA 19107  
zak@che.udel.edu

Francis J. Doyle III  
Department of Chemical  
Engineering  
University of California  
Santa Barbara, CA 93106  
doyle@engineering.ucsb.edu

James S. Schwaber  
Daniel Baugh Institute for  
Functional Genomics  
and Computational Biology  
Thomas Jefferson University  
Philadelphia, PA 19107  
james.schwaber@mail.tju.edu

## 1. INTRODUCTION

It is anticipated that regulatory networks governing diverse cellular behavior will be discovered through the analysis of functional genomic data. Computational models may play an essential role in this task, as they can generate *in silico* data for validating analysis methods [7, 6], and their properties may be analyzed to gain system-level insights. In the present study, local identifiability analysis is applied to a small computational model of a genetic regulatory network. The objective is to determine whether or not it is theoretically possible to uncover the network architecture using microarray (gene expression) data. It was observed that identifying the network architecture may only be possible when a rich microarray time course is coupled with information that specifies which transcription factors bind to which genes.

## 2. COMPUTATIONAL MODEL

The model used in the present study (Figure 1) is from a larger model that has been reported previously [7]. It is based on agonist-induced down-regulation of a steroid receptor [2]. A soluble ligand ( $Q$ ) diffuses through the plasma membrane and into the nucleus where it binds to and activates a transcription factor (the steroid receptor,  $E$ ), which then causes changes in the expression of target genes ( $F$ ). In this model, these changes in gene expression ultimately lead to down-regulation of the steroid receptor.

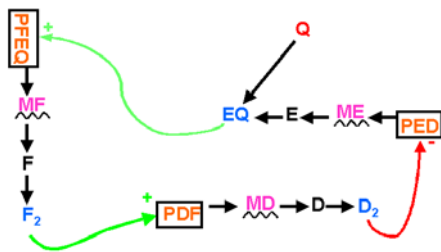


Figure 1: Model genetic regulatory network

For the present analysis, this system is represented by a set of ordinary differential equations (ODEs). Transcriptional regulation is modeled in the manner of Barkai and Leibler (2000) [1]. Equation 1 shows the equations pertaining to  $D$ : promoter ( $PDF$ ) binding by a transcription factor ( $F_2$ ), transcription, translation of the transcript ( $MD$ ), dimerization of the protein ( $D$ ), and further promoter

binding. The other components are described similarly.

$$\begin{aligned} [P\dot{D}F] &= -k_{PDF}[PDF][F_2] + k_{UPDF}[F_2PDF] \\ [F_2\dot{P}DF] &= k_{PDF}[PDF][F_2] - k_{UPDF}[F_2PDF] \\ [\dot{M}D] &= k_{RPDF}[PDF] + k_{RFPDF}[F_2PDF] - k_{dMD}[MD] \\ [\dot{D}] &= k_{TD}[MD] - 2k_{D2}[D]^2 + 2k_{UD2}[D_2] - k_{dD}[D] \\ [\dot{D}_2] &= k_{D2}[D]^2 - k_{UD2}[D_2] - k_{dD2}[D_2] \\ &\quad - k_{PED}[PED][D_2] + k_{UPED}[D_2PED] \end{aligned} \quad (1)$$

The overall model consists of 13 states and 31 parameters. Parameter values were taken from similar systems in the literature. Sample time courses of the transcripts ( $MD$ ,  $ME$ , and  $MF$ ) in response to ligand inputs are shown in Figure 2.

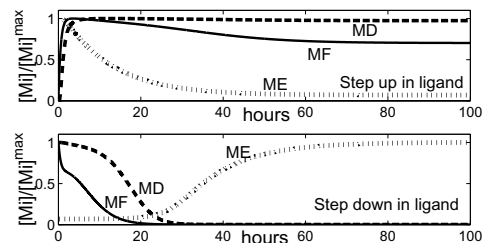


Figure 2: Expression profiles in response to a step up (top) and step down (bottom) in ligand ( $Q$ ) concentration.

In the present study, the problem of identifying the architecture of the system in Figure 1 from expression profiles of  $MD$ ,  $ME$ , and  $MF$  for cases with and without *localization information* are considered. Localization information specifies which transcription factors bind to which promoters (for example [5]). When there is localization information available, the identification of the system architecture involves identifying the transcription rate constants for the bound and unbound promoters for each gene ( $k_{RPDF}$  and  $k_{RFPDF}$  for gene  $D$ , for example), six parameters in total. In the case without localization information, identifying the architecture involves identifying the binding constants for each transcription factor for each gene as well as the transcription rate constants, giving a total of 27 parameters ( $3 \times 3 + 6 \times 3$ ), of which only six are non-zero. The difference is illustrated in Figure 3, where each  $X$  corresponds to 3 parameters (1 promoter binding and 2 transcription rate constants) that must be identified.



Figure 3: Possible transcription factor–gene interactions in model, demonstrating the constraining effect of localization. (a) Localization is not known. (b) Localization is known.

### 3. LOCAL IDENTIFIABILITY ANALYSIS

A model parameter is locally identifiable if a set of ideal measurements determines a finite set of values for that parameter [4]. Local identifiability is weaker, but more computationally feasible, than global identifiability, which requires unique determination of parameter values. If a parameter is not locally or globally identifiable, its value can only be determined as a combination of other parameters. For this reason, identifiability analysis is a key component of parameter estimation. Following Jacquez and Perry (1990), local identifiability analysis involves first determining the  $N_s \times N_p$  sensitivity matrix,  $S(t)$  for the ODE system of  $N_s$  states ( $x$ ) and  $N_p$  parameters ( $p$ ). It is rare to measure all of the dynamic states ( $x$ ) in an experiment, and thus the sensitivities of the measurements,  $S'(t)$ , may be given by  $S'(t) = CS(t)$ , where the measurements ( $y$ ) are related to the states by  $y(t) = Cx(t)$ . Values of  $S'$  at several different times may be stacked to give  $G$ :

$$G \triangleq [S'(t_1), S'(t_2), \dots]^T \quad (2)$$

The system is locally identifiable if  $\det(G^T G) \neq 0$ . When  $\det(G^T G) = 0$ , putting it into reduced row echelon form will reveal rows that contain single nonzero elements. The column indices of these elements indicate locally identifiable parameters. Rows containing multiple nonzero elements indicate parameters that are not locally identifiable [4]. Note that when this analysis is applied to nonlinear systems, absence of local identifiability does not guarantee absence of local identifiability for the nonlinear system, while the reverse is true for local identifiability [4].

### 4. RESULTS AND CONCLUSIONS

The local identifiability analysis was performed for cases where microarray data was used alone or with localization information for three experimental designs of ligand step up, step down, and combined steps. Results are shown in Figure 4. If microarray data is used alone, the parameters that determine the network architecture are not locally identifiable. Inclusion of localization information allows the network architecture parameters for gene  $E$  to become locally identifiable for either of the steps (step down not shown). Only when both steps are used together with localization information do all of the network architecture parameters become identifiable. For this case, there are also many more types of parameters that become locally identifiable.

Even with localization information, the network architecture was identifiable only when both steps were used together. These results demonstrate the importance of *rich* experimental data. The microarray data that may be acquired in the near future, however, may not be rich enough for determination of network architecture. Even with rich data, in the absence of localization information, the parameters that determine network architecture were not identifiable. This motivates combining microarray data with other information that constrains possible interactions, and is consistent with previous studies that demonstrated that inclusion of localization information improved determination of regulatory network architectures [7, 3].

(a) STEP UP ALONE

PARAMETER TYPE	NUMBER OF IDENTIFIABLE PARAMS.	
	microarray alone	microarray & localization
promoter binding	4/9	1/3
transcription	0/18	4/6
transcript degradation	0/3	3/3
protein degradation	0/3	1/3
dimer degradation	0/3	1/3

(b) STEP UP & STEP DOWN

PARAMETER TYPE	NUMBER OF IDENTIFIABLE PARAMS.	
	microarray alone	microarray & localization
promoter binding	4/9	3/3
transcription	0/18	6/6
promoter unbinding	0/9	2/3
transcript degradation	0/3	3/3
protein degradation	0/3	3/3
dimer degradation	0/3	3/3
dimerization	0/3	2/3
translation	0/3	2/3
ligand degradation	1/1	1/1

Figure 4: Localization and rich data are necessary to determine network architecture. Locally identifiable parameters when microarray data is used alone or with localization information. (a) Step up in ligand alone. (b) Step up and step down used together. Emphasized parameters are those necessary for determining network architecture.

Future work will involve the application of local identifiability analysis to larger, more complex regulatory network models [7] and to pursue more rigorous nonlinear identifiability methods.

### 5. ACKNOWLEDGMENTS

We acknowledge Dr. Joerg Stelling and Kapil Gadkar for discussions regarding sensitivity and identifiability analyses, respectively. We thank NIH NIGMS, NIH HLB, and DARPA BioCOMP for funding. DEZ also acknowledges NIH training grant NIAAA5T32AA07463–15 for funding.

### 6. REFERENCES

- [1] N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, 403:267–268, 2000.
- [2] A. H. Brivanlou and J. E. Darnell. Signal transduction and the control of gene expression. *Science*, 295:813–818, February 2002.
- [3] A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, 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.
- [4] J. A. Jacquez and T. Perry. Parameter estimation: local identifiability of parameters. *Am. J. Physiol.*, 258(4):E727–E736, 1990.
- [5] 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.
- [6] V. A. Smith, E. D. Jarvis, and A. J. Hartemink. Evaluating functional network inference using simulations of complex biological systems. *Bioinformatics*, In press.
- [7] D. E. Zak, F. J. Doyle, 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.