Modeling and Identification of the Gene Regulatory Network Describing the Liver Response to Corticosteroids

Gregory M. Miller, Daniel E. Zak*, James S. Schwaber, Babatunde A. Ogunnaike*

Efforts to model and identify gene regulatory networks have typically involved the consideration of every gene potentially regulating every other gene (Brazhnik et al., 2002). For most organisms of interest, the genome is so large that determination of the gene regulatory network on this basis becomes intractable due to the enormous number of parameters that must be estimated from limited experimental data.

Recently, we have described a structured framework for modeling and identification of gene regulatory networks (Zak et al., in review). The structured approach divides a cell into two subcellular regions: nuclear and cytoplasmic. The cytoplasmic model seeks to describe the activity of the transcription factors in response to extracellular signals and gene expression; the nuclear model describes gene activation as a function of transcription factor interactions. Such subcellular organization for the model may be obtained from prior knowledge of gene functions and intracellular signaling pathways.

The nuclear model may be further organized into the nuclear connectivity network, which specifies which transcription factors regulate which genes. This network may be determined experimentally, or by combining clustering approaches with gene expression patterns, promoter sequences, and algorithms or databases for identifying transcriptional regulatory elements (Vadigepalli et al., 2003), as has been done in previous studies (Tavazoie et al., 1999). Instead of identifying network structure and model parameters simultaneously as in unstructured approaches, structured model identification involves the determination of only those parameters that quantify the functional interactions between network components. Thus, by largely defining network structure through the inclusion of available biological information, the structured modeling approach is well suited for constructing models from microarray data that are often limited in quantity and quality (Nadon and Shoemaker, 2002).

The present work applies the structured approach to model the response of rat hepatocytes to glucocorticoids as a case study for a mammalian system. We selected this system as a test-bed because of (1) its physiological importance as an anti-inflammatory drug (Jin et al., 2003), (2) the availability of a high quality cytoplasmic model for the initial system response (Ramakrishnan et al., 2002), and (3) the availability of *in vivo* microarray time course data that are of sufficient quality to support the modeling approach we have proposed (Almon et al., 2003).

Thus far our work with this system has centered on the development of its nuclear connectivity structure. One key step in this process involves the grouping of genes on the basis of similarity in their expression profiles, and numerous clustering approaches to accomplish this have been described (Sherlock, 2000). An alternative clustering method has been developed that accounts for the essential dynamics of gene transcription (Sasik et al., 2002). This method assumes that the expression time-course for any regulated gene is the response of a first-order dynamic system to a finite-width pulse, where the degradation constant for each gene has been assumed to be invariant and transcription has been idealized

as a pulse input occurring between a particular onset and cessation times. Parameters in this first order finite width pulse response (FOFWPR) model may be estimated for each gene from its expression profile via least-squares minimization. Genes are then grouped on the basis of similarity in their estimated parameters. Since this method accounts for basic transcriptional dynamics, and reduces each expression profile to four characteristic parameters, we have employed it in the present work.

Only a subset of all genes in the dataset was adequately described by the FOFWPR model, as would be expected given that not all genes in the system will be responsive to glucocorticoids. Preliminary promoter analysis of the gene groups suggests that the regulation underlying the liver transcriptional response to glucocorticoids involves a complex interplay between several transcription factors. For example, we found that the promoters of the 150 upregulated genes that best fit the FOFWPR model were significantly ($p < 0.01$) enriched for binding sites for the transcription factors Ikaros and CRE-BP1 (ATF2). Antagonistic and activating interactions between glucocorticoids and these transcription factors have been reported previously (Newell et al., 1994; Wargnier et al., 1998). Our future work will involve further development of the nuclear connectivity structure for this system through consideration of additional clustering approaches. We will then integrate the nuclear connectivity with the existing cytoplasmic model for this system and the gene expression profiles for the relevant transcription factors, leading to a predictive, integrative model that describes the complex transcriptional regulation underlying this physiologically important response.

The focus of our presentation will be on the formulated model of the gene regulatory network describing the liver response to corticosteroids. Emphasis will be placed on how the model was constructed, and the assumptions made during its synthesis. In particular, the nuclear connectivity structure developed for this system will be described in detail, and the efficacy of the FOFWPR model as a technique for clustering will be reported.

Almon, R.R., et al. (2003). *Pharmacogenomics*, **4**(6), 791-9.

Brazhnik, P., et al. (2002). *Trends Biotechnol.*, **20**(11), 467-72.

Jin, J.Y., et al. (2003). *J. Pharmacol. Exp. Ther.*, **307**(1), 93-109.

Nadon, R., & Shoemaker, J. (2002). *Trends Genet.*, **18**(5), 65-71.

Newell, C.L., et al. (1994). *J. Leukoc. Biol.*, **56**(1), 27-35.

Ramakrishnan, R., et al. (2002). *J. Pharmacokinet. Pharmacodyn.*, **29**(1), 1-24.

Sasik, R., et al. (2002). *Bioinformatics*, **18**(1), 61-6.

Sherlock, G. (2000). *Curr. Opin. Immunol*, **12**(2), 201-5.

Tavazoie, S., et al. (1999). *Nat. Genet.*, **22**(3), 281-5.

Vadigepalli, R., et al. (2003). *OMICS*, **7**(3), 235-52.

Wargnier, A., et al. (1998). *J. Biol. Chem.*, **273**(52), 35326-31.

Zak, D.E., et al. *Computers and Chemical Engineering*, In Review.