A system for generating transcription regulatory networks with combinatorial control of transcription

Sushmita Roy, Margaret Werner-Washburne and Terran Lane

1Department of Computer Science and 2Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA

Received on January 21, 2008; revised on March 13, 2008; accepted on April 4, 2008

Advance Access publication April 8, 2008

ABSTRACT

Summary: We have developed a new software system, REgulatory Network generator with COmbinatorial control (RENO), for automatic generation of differential equations describing pre-transcriptional combinatorial networks in artificial regulatory networks. RENCO has the following benefits: (a) it explicitly models protein–protein interactions among transcription factors, (b) it captures combinatorial control of transcription factors on target genes and (c) it produces output in Systems Biology Markup Language (SBML) format, which allows these equations to be directly imported into existing simulators. Explicit modeling of the protein interactions allows RENCO to incorporate greater mechanistic detail of the transcription machinery compared to existing models and can provide a better assessment of algorithms for regulatory network inference.

Availability: RENCO is a C++ command line program, available at http://sourceforge.net/projects/renco/

Contact: terran@cs.unm.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

With the increasing availability of genome-scale data, a plethora of algorithms are being developed to infer regulatory networks. Examples of such algorithms include Bayesian networks, ARACNE (Bansal et al., 2007). Because of the absence of “ground truth” of regulatory network topology, these algorithms are evaluated on artificial networks generated via network simulators (Kurata et al., 2003; Margolin et al., 2005; Mendes et al., 2003; Schilstra and Bolouri, 2002).

Since gene regulation is a dynamic process, existing network simulations employ systems of ordinary differential equations (ODEs) that describe the kinetics of mRNA and protein concentrations as a function of time. Some approaches construct highly detailed models, but require large amounts of user-specified information (Kurata et al., 2003; Schilstra and Bolouri, 2002). Other approaches generate large networks but use simpler models by making the mRNA concentration of target genes dependent upon mRNA concentration, rather than on protein concentration of transcription factors (Mendes et al., 2003). In real biological systems, protein expression does not correlate with gene expression, especially at steady state, due to different translation and degradation rates (Belle et al., 2006). These approaches also do not model protein interactions edges and, therefore, combinatorics resulting from these interactions.

We describe a regulatory network generator, RENCO, that models genes and proteins as separate entities, incorporates protein–protein interactions among the transcription factor proteins, and generates ODEs that explicitly capture the combinatorial control of transcription factors. RENCO accepts either pre-specified network topologies or gene counts, in which case it generates a network topology. The network topology is used to generate ODEs that capture combinatorial control among transcription factor proteins. The output from RENCO is in SBML format, compatible with existing simulators such as Copasi (Hoops et al., 2006) and RANGE (Long and Roth, 2007). Time-series and steady-state expression data produced from the ODEs from our generator can be leveraged for comparative analysis of different network inference algorithms.

2 TRANSCRIPTIONAL REGULATORY NETWORK GENERATOR

RENO works in two steps: (a) generate/read the network topology and (b) generate the ODEs specifying the transcription kinetics (see RENCO manual for details). For (a) proteins are connected to each other via a scale-free network (Albert and Barabasi, 2000), and to genes via a network with exponential degree distribution (Maslov and Sneppen, 2005).

2.1 Modeling combinatorial control of gene regulation

We model combinatorial control by first identifying the set of cliques, $C$, up to a maximum of size $t$ in the protein interaction network. Each clique represents a protein complex that must function together to produce the desired target regulation. A target gene, $g_t$, is regulated by $k$ randomly selected such cliques, where $k$ is the indegree of the gene. These $k$ cliques regulate $g_t$ by binding in different combinations, thus exercising combinatorial gene regulation. We refer to the set of cliques in a combination as a transcription factor complex (TFC). At any time there can be several such TFCs regulating $g_t$. The mRNA concentration of a target gene is, therefore, a function of three...
types of regulation: within-clique, within-complex and across-
complex regulation. Within-clique regulation captures the
contribution of one clique on a target gene. The within-
complex regulation captures the combined contribution of all
cliques in one TFC. Finally, the across-complex regulation
specifies the combined contribution of different TFCs.

We now introduce the notation for ODEs generated by
RENSO. \( M_i(t) \) and \( P_i(t) \) denote the mRNA and protein
concentrations, respectively, of gene \( g_i \), at time \( t \). \( V_i^M \) and \( V_i^P \) denote the rate constants of mRNA synthesis and degradation
of \( g_i \), \( v_i^M \) and \( v_i^P \) denote the rate constants of protein synthesis and
degradation. \( C_{ij} \) and \( T_{ij} \) denote a protein clique and a TFC,
respectively, associated with \( g_i \). \( Q_i \) denotes the set of TFCs
associated with \( g_i \). \( X_{ij} \), \( Y_{ij} \) and \( S_i \) specify the within-clique,
within-complex and across-complex regulation on \( g_i \).

Based on existing work (Mendes \textit{et al.}, 2003; Schilstra and
Bolouri, 2002), the rate of change of mRNA concentration
is the difference of synthesis and degradation of \( M_i \):
\[ \frac{dM_i(t)}{dt} = V_i^M S_i - v_i^M M_i(t). \]
Similarly for protein concentration,
\[ \frac{dP_i(t)}{dt} = V_i^P M_i(t) - v_i^P P_i(t). \]

The across-complex regulation, \( S_i \), is a weighted sum of
correlations from \( |Q_i| \) TFCs: \( S_i = \sum_{q=1}^{Q_i} w_q Y_{ij} \), where \( w_q \)
denotes the TFC weight. The sum models ‘or’ behavior of the
different TFCs because all TFCs need not be active simulta-
neously. The within-complex regulation, \( Y_{ij} \) is a product
of within-clique actions in the TFC \( T_{ij} \), \( Y_{ij} = \prod_{i=1}^{C_i} X_{ij} \).
The product models ‘and’ behavior of a single TFC because all
proteins within a TFC must be active at the same time. Finally,
the cliques per gene \( C_{ij} \) are randomly assigned activating or
repressing roles on \( g_i \). If \( C_{ij} \) is activating,
\[ X_{ij} = \frac{\prod_{p=1}^{C_i} P_p(t)}{\prod_{j=1}^{C_i} K_{a_{ij}} + \prod_{j=1}^{C_i} P_p(t)} \]
otherwise,
\[ X_{ij} = \frac{\prod_{j=1}^{C_i} K_{i_{jp}}}{\prod_{j=1}^{C_i} K_{a_{ij}} + \prod_{j=1}^{C_i} P_p(t)} \]
\( K_{a_{ij}} \) and \( K_{i_{jp}} \) are equilibrium dissociation constants of the \( p \)
th activator or repressor of \( g_i \). All degradation, synthesis and
dissociation constants are initialized uniformly at random from
\([0.01, V_{\text{max}}]\), where \( V_{\text{max}} \) is user specified.

\section{EXAMPLE NETWORK}

We used RENCO to analyze: (a) mRNA and protein steady-
state measurements and (b) combinatorial gene regulation,
in a small example network (Supplementary Material has
details).

\subsection{Importance of modeling protein expression}

The example network has five genes and five proteins (Fig. 1a).
The gene \( G_4 \) is regulated via different combinations of the
cliques \( \{P_2\}, \{P_0, P_1\} \). We find that the wild-type time courses of
individual mRNA expressions are correlated with corresponding
proteins (Fig. 1b and c). But because different genes and
proteins have different degradation and synthesis rate constants,
the mRNA population as a whole does not correlate with the
protein population (Spearman’s correlation = 0.3). Because of
the dissimilarity in the steady-state mRNA and protein expres-
sion populations, genes appearing to be differentially expressed
at the mRNA level may not be differentially expressed at the
protein level. This highlights the importance of modeling
mRNA and protein expression as separate entities in the
network.

\subsection{Combinatorics of gene regulation}

We analyzed combinatorial control in our network by generat-
ing the \( G_4 \) time course under different knock-out combinations of
the \( G_4 \) activators, \( P_0 \) and \( P_1 \) (Fig. 2). Because all the
regulators are activating, \( G_4 \) is downregulated here compared to
wild-type. We note that each knock out combination yields
different time courses. In particular, knocking out either \( G_0 \) or
\( G_1 \) in combination with \( G_2 \) is sufficient to drive the \( G_4 \) expression
to 0. This phenomenon is because of the clique, \( P_0, P_1 \).
This illustrates a possible combinatorial regulation process to produce a range of expression dynamics using a few transcription factors.

4 CONCLUSION
We have described RENCO, a generator for artificial regulatory networks and their ODEs. RENCO models the transcriptional machinery more faithfully by explicitly capturing protein interactions and provides a good testbed for network structure inference algorithms.

ACKNOWLEDGEMENTS

Funding: This work was supported by an HHMI-NIH/NIBIB grant (56005678), an NSF (MCB0645854) grants to M.W.W., and an NIMH grant (1R01MH076282-01) to T.L.

Conflict of Interest: none declared.

REFERENCES
Albert,R. and Barabasi,A.-L. (2000) Topology of evolving networks: local events and universality. Phys. Rev. Lett., 85, 5234–5237.
Bansal,M. et al. (2007) How to infer gene networks from expression profile. Mol. Syst. Biol., 3.
Belle,A. et al. (2006) Quantification of protein half-lives in the budding yeast proteome. PNAS, 103, 13004–13009.
Hoops,S. et al. (2006) Copasi – a complex pathway simulator. Bioinformatics, 22, 3067–3074.
Kurata,H. et al. (2003) CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle. Nucl. Acids Res., 31, 4071–4084.
Long,J. and Roth,M. (2007) Synthetic microarray data generation with range and nemo. Bioinformatics, 24, 132–134.
Margolin,A. et al. (2005) Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics, 7 (Suppl 1): S7.
Maslov,S. and Sneppen,K. (2005) Computational architecture of the yeast regulatory network. Physical Biology, 2, s94–s100.
Mendes,P. et al. (2003) Artificial gene networks for objective comparison of analysis algorithms. Bioinformatics, 19, 122–129.
Schilstra,M.J. and Bolouri,H. (2002) The Logic of Life. In Proceedings of 3rd International Conference on Systems Biology (ICSB). Karolinska Institutet, Stockholm, Sweden.