# Multi-objective Model Optimization for Inferring Gene Regulatory Networks

Christian Spieth, Felix Streichert, Nora Speer, and Andreas Zell

Centre for Bioinformatics Tübingen (ZBIT), University of Tübingen,
Sand 1, D-72076 Tübingen, Germany
spieth@informatik.uni-tuebingen.de
http://www-ra.informatik.uni-tuebingen.de

**Abstract.** With the invention of microarray technology, researchers are able to measure the expression levels of ten thousands of genes in parallel at various time points of a biological process. The investigation of gene regulatory networks has become one of the major topics in Systems Biology. In this paper we address the problem of finding gene regulatory networks from experimental DNA microarray data. We suggest to use a multi-objective evolutionary algorithm to identify the parameters of a non-linear system given by the observed data. Currently, only limited information on gene regulatory pathways is available in Systems Biology. Not only the actual parameters of the examined system are unknown, also the connectivity of the components is a priori not known. However, this number is crucial for the inference process. Therefore, we propose a method, which uses the connectivity as an optimization objective in addition to the data dissimilarity (relative standard error - RSE) between experimental and simulated data.

## 1 Introduction

Gene regulatory networks (GRNs) represent the dependencies of the different actors in a cell operating at the genetic level. They dynamically determine the level of gene expression for each gene in the genome by controlling whether a gene will be transcribed into RNA or not. A simple GRN consists of one or more input signalling pathways, several target genes, and the RNA and proteins produced from those target genes. In addition, such networks often include dynamic feedback loops that provide further network regulation activities and output. In order to understand the underlying structures of activities and interactions of intra-cellular processes one has to understand the dependencies of gene products and their impact on the expression of other genes. Therefore, finding a GRN for a specific biological process would explain this process from a logical point of view, thus explaining many diseases.

Therefore, the model reconstruction of gene regulatory networks has become one of the major topics in bioinformatics. However, the huge number of system components requires a large amount of experimental data to infer genome-wide networks. Recently, DNA microarrays have become one of the major tools in the

C. A. Coello Coello et al. (Eds.): EMO 2005, LNCS 3410, pp. 607–620, 2005.

research area of microbiology. This technology enables researchers to monitor the activities of thousands of genes in parallel and can therefore be used as a powerful tool to understand the regulatory mechanisms of gene expression in a cell. With this technique, cells can be studied under several conditions such as medical treatment or different environmental influences.

Microarray experiments often result in time series of measured values indicating the activation level of each tested gene in a genome. These data series can then be used to examine the reactions of the cell to external stimuli. A model would enable biologists to predict the reactions of intracellular signalling processes. To re-engineer or infer the regulatory processes computationally from these experimental data sets, one has to find a model that is able to produce the same time series data as the experiments. The idea is then that the model reflects the true system dependencies, i.e. the dependencies of the components of the regulatory system.

Several approaches have been made to address this problem. Many of them are only relying on the distance between the experimental data and the simulated data coming from the mathematical model, but the biological plausibility of the system is almost always neglected. And although Biologist know, that regulatory systems are sparse, i.e. one gene relies on average on a small number of other genes, this fact can be found only in some publications.

In this paper we propose a methodology for reverse engineering sets of time series data obtained by artificial expression analysis by combining two objectives into a multi-objective optimization problem. The first objective is the dissimilarity between the experimental and the simulated data (RSE). The second objective is the connectivity of the system. Both objectives are to be minimized to gain a system, which fits the data and at the same time is only sparsely connected and therefore biological plausible. With this approach, we systematically examine the impact of the connectivity of the regulatory network on the overall inference process.

The remainder of this paper is structured as follows. Section 2 of this paper presents an overview over related work and lists associated publications. Detailed description of our proposed method will be given in section 3 and example applications will be shown in section 4. Finally, conclusions and an outlook on future research will be covered by section 5.

## 2 Related Work

Researchers are interested in understanding the mechanisms of gene regulatory processes and therefore in inferring the underlying networks. The following section briefly describes the work that has been done in this area.

The earliest models to simulate regulatory systems found in the literature are boolean or random boolean networks (RBN) [8]. In boolean networks gene expression levels can be in one of two states: either 1 (on) or 0 (off). The quantitative level of expression is not considered. Two examples of algorithms for inferring GRNs with boolean networks are given by Akutsu [1] and the RE-