Bio-mimetic Evolutionary Reverse Engineering of Genetic Regulatory Networks

Daniel Marbach, Claudio Mattiussi, and Dario Floreano

Ecole Polytechnique Fédérale de Lausanne (EPFL),
Laboratory of Intelligent Systems,
CH-1015 Lausanne, Switzerland
{Daniel.Marbach,Claudio.Mattiussi,Dario.Floreano}@epfl.ch
http://lis.epfl.ch

Abstract. The effective reverse engineering of biochemical networks is one of the great challenges of systems biology. The contribution of this paper is two-fold: 1) We introduce a new method for reverse engineering genetic regulatory networks from gene expression data; 2) We demonstrate how nonlinear gene networks can be inferred from steady-state data alone. The reverse engineering method is based on an evolutionary algorithm that employs a novel representation called Analog Genetic Encoding (AGE), which is inspired from the natural encoding of genetic regulatory networks. AGE can be used with biologically plausible, nonlinear gene models where analytical approaches or local gradient based optimisation methods often fail. Recently there has been increasing interest in reverse engineering *linear* gene networks from steady-state data. Here we demonstrate how more accurate *nonlinear dynamical models* can also be inferred from steady-state data alone.

Keywords: Systems Biology, Gene Networks, Reverse Engineering, Steady-State Data, Genetic Algorithm, Analog Genetic Encoding (AGE).

1 Introduction

Genetic regulatory networks perform fundamental information processing and control mechanisms in the cell. Regulatory genes code for proteins that enhance or inhibit the expression of other regulatory and/or non-regulatory genes, thereby forming a complex web of interactions (Fig. 1a). Inference and simulation of gene networks may contribute substantially to our biological knowledge in the post-genomic era. Practical applications may have a strong impact on biotech and pharmaceutical industries, potentially setting the stage for rational redesign of living systems and predictive, model-based drug design [1]. Technologies to assay gene expression levels in terms of mRNA concentrations are advancing at a fast pace. Using oligonucleotide chips or quantitative PCR for instance, it is possible to probe a set of genes of interest that are part of an uncharacterized gene network (henceforth known as *target network*) under different conditions. The goal of reverse engineering is inferring the target gene regulatory network from this experimental data.

The choice of a suitable reverse engineering method depends on the type of model used to describe the target network. Here we focus on models that represent

E. Marchiori, J.H. Moore, and J.C. Rajapakse (Eds.): EvoBIO 2007, LNCS 4447, pp. 155–165, 2007.

a genetic regulatory network as a dynamical system described by a system of ordinary differential equations. The linear model [1,2,3,4,5], which is based on a first-order approximation of gene expression dynamics, is by far the most widely used gene network model. Its main advantage is that reverse engineering can be tackled analytically using standard techniques of system identification [1,2,3,4,5,6]. However, gene regulation is known to be strongly nonlinear. Hence, the linearization is generally only valid in a small regime, i.e., close to a specific steady-state [1,3,5]. This implies that valuable data from perturbation experiments with a strong effect on the network (e.g., gene knockouts) cannot be used because they fall outside the valid regime of the first-order approximation [1,5]. Furthermore, the inferred linear model is unlikely to correctly predict network response under strong perturbations [1,5], as can be expected in disease for instance.

As both quantity and quality of experimental data improve, we can aim at a more biologically plausible, faithful reconstruction of the target network. This requires the conception of adequate inference methods that can handle complex, nonlinear gene models, where analytical approaches and local gradient based optimisation often fail. In this paper we propose a bio-mimetic approach based on artificial evolution [7] using Analog Genetic Encoding (AGE) [8,9], an artificial genetic representation that has already proven its merits in benchmark problems in the fields of analog electronic circuits [8,9] and artificial neural networks [10]. Unlike other reverse engineering algorithms based on global optimisation techniques such as simulated annealing [11,12] or conventional evolutionary algorithms [4,13,14,15], AGE allows simultaneous inference of model structure and numerical parameter values. Furthermore, AGE mimics the evolutionary process of incremental complexification that natural gene regulatory networks are subjected to.

In the past, most reverse engineering studies used time-series gene expression data. However, time-series data are more difficult to obtain experimentally than steady-state data and their information content is lower (samples in a time-series are not independent). Indeed, there has been a recent trend towards approaches based on steady-state perturbation data [1,3,16,17]. These studies use analytical approaches based on first-order approximations. Here we demonstrate for the first time – to the best of our knowledge – how *nonlinear* models (network structure and parameters) can be inferred from steady-state data alone.

2 Evolutionary Reverse Engineering with AGE

The first step in the reverse engineering process generally consists in the choice of a gene network model type (e.g. the sigmoid model introduced in Sect. 3). AGE is not constrained to a specific model type, but can be used with a large class of nonlinear models termed *analog networks* [9]. An analog network is composed of a collection of devices connected by links of different strengths. Here, devices are genes and links correspond to regulatory interactions[1]. Without limiting

[1] In this paper we consider the simplest case where all devices are of the same type, but AGE can also handle heterogeneous networks [9]. We plan to use several device types in the future for more complex models of gene-protein networks, for instance.