BioSystems 116 (2014) 36–42

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Evolving Boolean regulatory networks with epigenetic control

Larry Bull *

Department of Computer Science & Creative Technologies, University of the West of England, Bristol BS16 1QY, UK

ARTICLE INFO

Article history:
Received 20 June 2013
Received in revised form 15 October 2013
Accepted 9 December 2013

Keywords:
Development
Dynamical systems
Methylation
Multicellularity
RBNK model

ABSTRACT

The significant role of epigenetic mechanisms within natural systems has become increasingly clear. This paper uses a recently presented abstract, tunable Boolean genetic regulatory network model to explore aspects of epigenetics. It is shown how dynamically controlling transcription via a DNA methylation-inspired mechanism can be selected for by simulated evolution under various single and multicellular scenarios. Further, it is shown that the effects of such control can be inherited without detriment to fitness.

1. Introduction

Epigenetics refers to cellular mechanisms that affect transcription without altering DNA sequences, e.g., see (Bird, 2007) for an overview. The two principal mechanisms are DNA methylation and histone modification. In the former case, a methyl group attaches to the base cytosine, or adenine in bacteria, typically causing a reduction in transcription activity in the area. In the latter case, changes in the shape of the proteins around which DNA wraps itself to form chromatin can alter the level of transcription in the area. In both cases, the change can be inherited.

With the aim of enabling the systematic exploration of artificial genetic regulatory network models (GRN), a simple approach to combining them with abstract fitness landscapes has recently been presented (Bull, 2012). More specifically, random Boolean networks (RBN) (Kauffman, 1969) were combined with the NK model of fitness landscapes (Kauffman and Levin, 1987). In the combined form – termed the RBNK model – a simple relationship between the states of N randomly assigned nodes within an RBN is assumed such that their value is used within a given NK fitness landscape of trait dependencies. The approach was also extended to enable consideration of multicellular scenarios using the related NKCS landscapes (Kauffman and Johnsen, 1992) – termed the RBNKCS model.

In this paper, RBNs are extended to include a simple form of epigenetic control. The selective advantage of the new mechanism is explored under various single and multicellular scenarios. Results indicate epigenetics is useful across a wide range of conditions. The paper is arranged are follows: the next section briefly reviews related work in the area and introduces the two basic models; Section 3 examines the extended RBNK model; and, Section 4 examines the extended RBNKCS model. Finally, all findings are discussed.

2. Background

2.1. Epigenetic computing

Whilst there is a growing body of work using artificial GRN within bio-inspired computing (e.g., see (Bull, 2012) for an overview), there are very few examples which consider epigenetic mechanisms explicitly. Note this is not the same as epigenetic robotics (e.g., see (Asada et al., 2009)). Tanav and Yuta (2003) included a histone modification-inspired scheme into a two-cell, rule-based representation where a development phase repeatedly alters one of the two cells. Periyasamy et al. (2008) presented an approach in which each individual in the evolving population is essentially viewed as a protein interacting with other proteins based upon various external and internal conditions, an architecture reminiscent of the Learning Classifier System (Holland, 1976). Turner et al. (2013) have recently augmented a GRN model with an epigenetic layer in the form of a set of binary masks over the genes, one mask per objective faced by the system: for a given task, the subset of genes defined in the corresponding mask are used to build the GRN. In this paper, a simple, abstract epigenetic mechanism is introduced into a well-known artificial GRN model which is an on-going, context dependent control process during the cell lifecycle.

It can be noted that, following (Maynard-Smith, 1990), a growing number of formal models of epigenetic mechanisms exist of natural systems (e.g., see (Geoghegan and Spencer, 2013)). A review

* Tel.: +44 0117 3283161.
E-mail address: Larry.Bull@uwe.ac.uk

0303-2647/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biosystems.2013.12.004

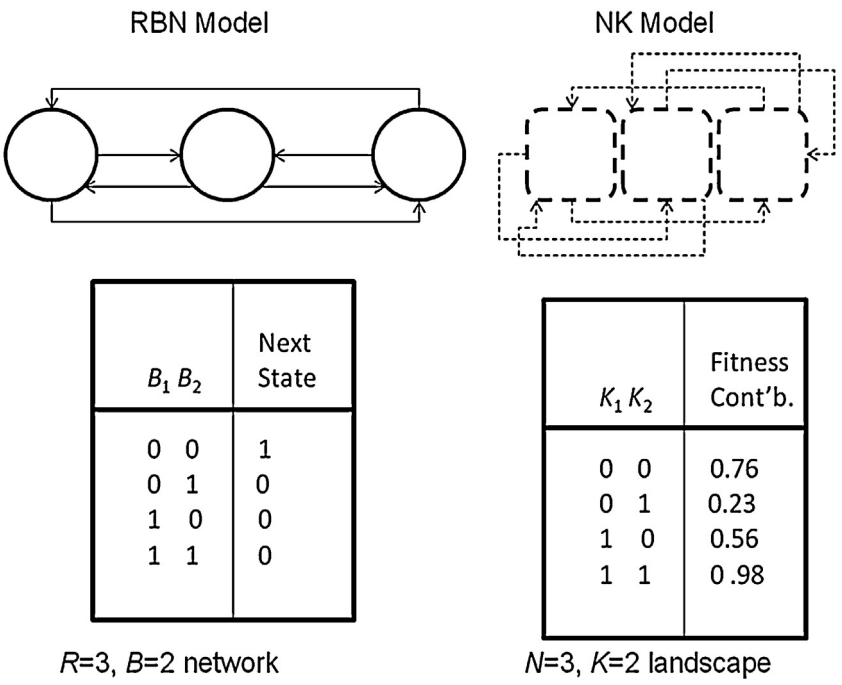

Fig. 1. Example traditional RBN (left) and NK (right) models. Both contain three genes mutually connected, with the state-transition/fitness-contribution table shown for one gene in each case.

of that literature is beyond the scope of this paper and the reader is referred to (Jablonka and Lamb, 2005) for an introduction.

2.2. The RBNK model

Within the traditional form of RBN, a network of R nodes, each with a randomly assigned Boolean update function and B directed connections randomly assigned from other nodes in the network, all update synchronously based upon the current state of those B nodes (Fig. 1). Hence those B nodes are seen to have a regulatory effect upon the given node, specified by the given Boolean function attributed to it. Since they have a finite number of possible states and they are deterministic, such networks eventually fall into an attractor. It is well-established that the value of B affects the emergent behaviour of RBN wherein attractors typically contain an increasing number of states with increasing B (see (Kauffman, 1993) for an overview). Three regimes of behaviour exist: ordered when $B = 1$, with attractors consisting of one or a few states; chaotic when $B \geq 3$, with a very large number of states per attractor; and, a critical regime around $B = 2$, where similar states lie on trajectories that tend to neither diverge nor converge (see (Derrida and Pomeau, 1986) for formal analysis). Note that traditionally the size of an RBN is labeled N, as opposed to R here, and the degree of node connectivity labeled K, as opposed to B here. The change is adopted due to the traditional use of the labels N and K in the NK model of fitness landscapes which are also used in this paper, as will be shown.

Kauffman and Levin (1987) introduced the NK model to allow the systematic study of various aspects of fitness landscapes (see (Kauffman, 1993) for an overview). In the standard NK model an individual is represented by a set of N (binary) genes or traits, each of which depends upon its own value and that of K randomly chosen others in the individual (Fig. 1). Thus increasing K, with respect to N, increases the epistasis. This increases the ruggedness of the fitness landscapes by increasing the number of fitness peaks. The NK model assumes all epistatic interactions are so complex that it is only appropriate to assign (uniform) random values to their effects

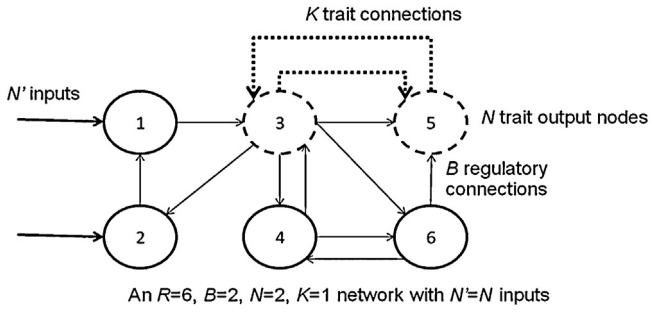

Fig. 2. Example RBNK model with an equal number of input and output nodes. Dashed lines and nodes indicate where the NK fitness landscape is embedded into the RBN model (refer to Fig. 1).

on fitness. Therefore for each of the possible K interactions, a table of $2^{(K+1)}$ fitnesses is created, with all entries in the range 0.0–1.0, such that there is one fitness value for each combination of traits. The fitness contribution of each trait is found from its individual table. These fitnesses are then summed and normalised by N to give the selective fitness of the individual. Exhaustive search of NK landscapes (Smith and Smith, 1999) suggests three general classes exist: unimodal when $K = 0$; uncorrelated, multi-peaked when $K > 3$; and, a critical regime around $0 < K < 4$, where multiple peaks are correlated.

As shown in Fig. 2, in the RBNK model N nodes (where $R \leq N < 0$) in the RBN are chosen as "outputs", i.e., their state determines fitness using the NK model. The combination of the RBN and NK model enables a systematic exploration of the relationship between phenotypic traits and the genetic regulatory network by which they are produced. It was previously shown how achievable fitness decreases with increasing B, how increasing N with respect to R decreases achievable fitness, and how R can be decreased without detriment to achievable fitness for low B (Bull, 2012). In this paper N phenotypic traits are attributed to randomly chosen nodes within the network of R genetic loci, with environmental inputs applied to the first N' loci (Figure 2); input nodes and trait/output nodes are