BioSystems 112 (2013) 56–62

Contents lists available at SciVerse ScienceDirect

BioSystems

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

The incorporation of epigenetics in artificial gene regulatory networks

Alexander P. Turner[a,d,*], Michael A. Lones[a,d], Luis A. Fuente[a,d], Susan Stepney[b,d], Leo S.D. Caves[c,d], Andy M. Tyrrell[a,d]

[a] Department of Electronics, University of York, Heslington, York YO10 5DD, UK
[b] Department of Computer Science, University of York, Heslington, York YO10 5GH, UK
[c] Department of Biology, University of York, Heslington, York YO10 5DD, UK
[d] York Centre for Complex Systems Analysis, University of York, Heslington, York YO10 5GH, UK

ARTICLE INFO

Keywords:
Artificial gene regulation
Epigenetics
Dynamical systems
Chaos control
Evolutionary algorithms

ABSTRACT

Artificial gene regulatory networks are computational models that draw inspiration from biological networks of gene regulation. Since their inception they have been used to infer knowledge about gene regulation and as methods of computation. These computational models have been shown to possess properties typically found in the biological world, such as robustness and self organisation. Recently, it has become apparent that epigenetic mechanisms play an important role in gene regulation. This paper describes a new model, the Artificial Epigenetic Regulatory Network (AERN) which builds upon existing models by adding an epigenetic control layer. Our results demonstrate that AERNs are more adept at controlling multiple opposing trajectories when applied to a chaos control task within a conservative dynamical system, suggesting that AERNs are an interesting area for further investigation.

1. Introduction

Gene regulatory networks are complex dynamical structures that underpin an organism's ability to control its internal environment (Turner, 2001). From a biological perspective, the study of gene regulation is of significant scientific importance because it determines cellular differentiation, which is responsible for the development of the different tissues and organs that underpin the structure of higher organisms (Latchman, 2005). From a computational perspective, gene networks are interesting because they are robust control structures, capable of dealing with environmental perturbations, whilst maintaining structure and order. Because of this, there has been significant interest in modelling gene regulatory networks *in silico* in order to capture these features e.g. (Lones et al., 2010; Quick et al., 2003). Due to the complexity of biological gene regulation, computational analogues are simplified. Research has shown that relatively simple networks, such as the random Boolean network, can exhibit emergent properties such as self-organisation and robustness and, in addition to this, can model real regulatory circuits (Kauffman, 1969; Kauffman et al., 2003).

However, current computational analogues of gene regulation fail to include models of one of the most pervasive methods of gene regulation in higher organisms, epigenetics. In this sense, the regulatory nature of these computational analogues may be limited in terms of their complexity and performance. In this paper, we define a computational representation of epigenetic information for the control of a complex dynamical system.

2. Gene regulation and epigenetics

A gene is a unit of hereditary information within a living organism, most commonly considered to be a region of DNA that specifies the primary structure of a protein. The genetic code is a biological blueprint that details which proteins can be produced, and ultimately, the phenotypic space within which the organism can exist. The lowest known threshold on the number of genes required to naturally facilitate life is that of *Mycoplasma genitalium*, which has approximately 470 genes (Fraser et al., 1995). Even in nature's most minimalist example of a gene regulatory network, 470 genes have to be coordinated in such a way as to maintain the optimum internal environment of the organism, highlighting that even the simplest of gene regulatory networks are inherently complex.

2.1. DNA methylation

One of the principal epigenetic mechanisms is DNA methylation, which refers to the addition of a methyl group to either the cytosine or adenine nucleobase in DNA (Fig. 1). It acts as an epigenetic

* Corresponding author at: Department of Electronics, University of York, Heslington, York YO10 5DD, UK.
E-mail addresses: apt503@york.ac.uk (A.P. Turner), michael.lones@york.ac.uk (M.A. Lones), laf509@york.ac.uk (L.A. Fuente), susan.stepney@cs.york.ac.uk (S. Stepney), leo.caves@york.ac.uk (L.S.D. Caves), andy.tyrrell@york.ac.uk (A.M. Tyrrell).

0303-2647/$ – see front matter © 2013 Published by Elsevier Ireland Ltd.
http://dx.doi.org/10.1016/j.biosystems.2013.03.013

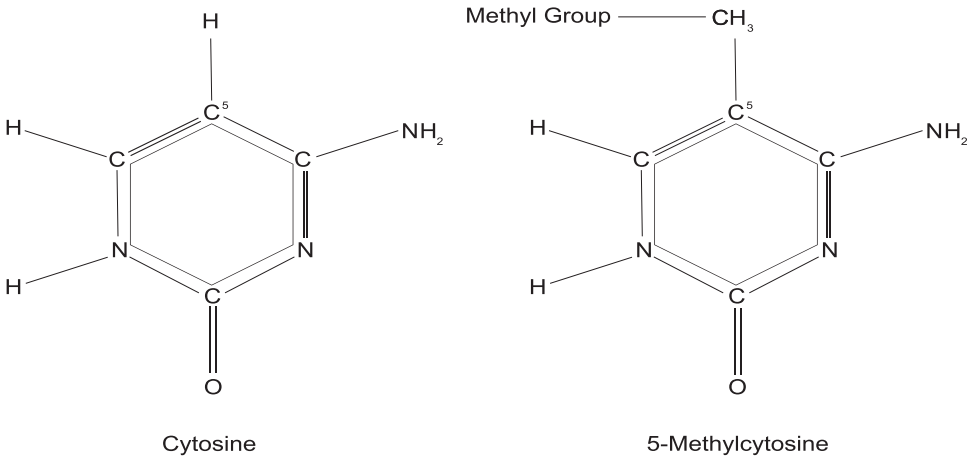

Fig. 1. DNA methylation. The attachment of a methyl group to a cytosine nucleobase has been shown to be able to regulate physiological processes (Bird, 2002; Hattman, 2005).

marker that can regulate many physiological processes (Bird, 2002; Hattman, 2005). In mammals, DNA methylation is most commonly present within CpG islands (Kim et al., 2009), locations in the DNA sequence with an abundance of cytosine guanine dinucleotides. In mice it has been shown that methylated cytosine bases account for approximately 1% of all the DNA bases within the genome, and therefore is representative of between 70 and 80% of all CpG dinucleotides in the genome (Kim et al., 2009).

2.2. Chromatin modifications

Especially in complex organisms, higher order structures such as chromatin have been shown to have significant influence over gene expression. Chromatin is a molecular complex consisting of a combination of DNA and proteins which is contained within the cell nucleus (Fig. 2). Chromatin has two functions. First, in the case of human cells, there are approximately 3400 Mb of DNA of approximately 2.3 m in length. Chromatin provides a structure for the condensation of DNA, so that it can be fully contained in a nucleus of approximately 6 μm (Alberts et al., 1994; Allis et al., 2007; Bushman, 2002). This is accomplished in multiple stages. The first is the wrapping of 145 base pairs of DNA in 1.67 toroidal superhelical turns around a histone octamer (Schones and Zhao, 2008). This forms the basis of the nucleosome, which moves through further levels of condensation into a solenoid fibre which ultimately folds into a chromosome (Fig. 2).

The second function of chromatin is to control access to the DNA, which in turn acts as an additional level of genetic control. This is because, in order for a gene to be expressed, the DNA has to be physically accessible by the transcriptional machinery within the cell. When it is tightly packed the structure of chromatin prevents this, allowing access to underlying DNA only when it is in its least condensed state.

2.3. Genes and epigenetic molecules

There is no clear separation between DNA methylation and chromatin modification in terms of gene expression. Research suggests that chromatin modification and DNA methylation are intrinsically linked (Jackson et al., 2002), and frequently, DNA methylation causes an underlying change to chromatin structure and gene expression (Jones and Takai, 2001). Generally, DNA methylation provides a more long term, stable effect on gene expression when compared to relatively short term reversible chromatin modifications (Cedar and Bergman, 2009). One of the more interesting aspects of epigenetics is that, in certain instances, epigenetic traits can be inherited by successive generations of cells, and sometimes organisms (Bird, 2007). In addition to this, epigenetic modifications can give the genetic code a relative genetic memory (Bird, 2002), which can then be used in such processes as cellular differentiation (Mohn and Schübeler, 2009).

It has been hypothesised that the characteristics of epigenetic molecules allows for a genetic plasticity which plays a major part in the development of complex phenotypes (Petronis, 2010). This plasticity creates the ability for organisms to express high levels of adaptability via utilisation of the dynamical reconfiguration afforded by epigenetic processes. In the following sections the idea of incorporating a level of artificial epigenetic information in an existing artificial gene regulatory network model is introduced.

3. Artificial genetic regulatory networks

Gene regulation in biology is a set of mechanisms that maintains homeostasis within an organism's internal environment. The aim of artificial gene regulation is to create a computational model of genetic behaviour that exhibits the useful and interesting properties of gene regulation in nature, namely self organisation, robustness and the expression of complex behaviours. The earliest example of this is the random Boolean network (RBN) (Kauffman, 1969). RBNs represent genes as Boolean expressions. These artificial genes are referred to as nodes. The network has a connectivity value, k, which specifies how many nodes influence a given node's expression level. The state of a node is defined by randomly initiated state transition rules. Upon execution, the network is iterated over a number of time steps, during which each node modifies its value depending upon its connectivity and its state transition rules. These networks demonstrate that with a k value of 2 or 3, distinct order and repetitive patterns can be generated. Moreover, for certain parameter ranges, the RBNs express high levels of robustness, maintaining relative order when exposed to external perturbations (Harvey and Bossomaier, 1997).

Subsequent models of gene regulation draw inspiration from the RBN. However there has been a shift towards continuous-valued models (Kumar, 2004; Lones et al., 2010), as they are computationally more flexible. In addition, it has been shown that these models can be applied to the control of complex systems (Lones et al., 2010).

3.1. Artificial epigenetic regulatory networks

This paper, a revised and extended version of (Turner et al., 2012) extends the model of artificial gene regulation (AGN)