Available online at www.sciencedirect.com
ScienceDirect
BioSystems 91 (2008) 216–230
BioSystems
www.elsevier.com/locate/biosystems

Rule-based computing system for microbial interactions and communications: Evolution in virtual bacterial populations

R. Gregory [a,*], V.A. Saunders [b], J.R. Saunders [c]

[a] *Department of Computer Science, Ashton Building, University of Liverpool, Liverpool L69 3BX, United Kingdom*
[b] *School of Biomolecular Sciences, Liverpool John Moores University, Liverpool L3 3AF, United Kingdom*
[c] *Microbiology and Genomics Division, School of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, United Kingdom*

Received 17 January 2007; received in revised form 23 July 2007; accepted 18 September 2007

Abstract

We have developed a novel rule-based computing system of microbial interactions and communications, referred to as COSMIC-Rules, for simulating evolutionary processes within populations of virtual bacteria. The model incorporates three levels: the bacterial genome, the bacterial cell and an environment inhabited by such cells. The virtual environment in COSMIC-Rules can contain multiple substances, whose relative toxicity or nutrient status is specified by the genome of the bacterium. Each substance may be distributed uniformly or in a user-defined manner. The organisms in COSMIC-Rules possess individually-defined physical locations, size, cell division status and genomes. Genes and/or gene systems are represented by abstractions that may summate sometimes complex phenotypes. Central to COSMIC-Rules is a simplified representation of bacterial species, each containing a functional genome including, where desired, extrachromosomal elements such as plasmids and/or bacteriophages. A widely applicable computer representation of biological recognition systems based on bit string matching is essential to the model. This representation permits, for example, the modelling of protein–protein interactions, receptor–ligand interactions and DNA–DNA transactions. COSMIC-Rules is intended to inform studies on bacterial adaptation and evolution, and to predict behaviour of populations of pathogenic bacteria and their viruses. The framework is constructed for parallel execution across a large number of machines and efficiently utilises a 64 processor development cluster. It will run on any Grid system and has successfully tested simulations with millions of bacteria, of multiple species and utilising multiple substrates. The model may be used for large-scale simulations where a genealogical record for individual organisms is required.

Keywords: Evolution; Individual-based-model (IbM); Bacteria; Genome; State machine

1. Introduction

Bacterial genetic systems have been used both as metaphors for developing computing architectures that may be of application to biology and other disciplines (e.g. Devine et al., 1996; Ginovart et al., 2002; Gregory et al., 2004a), and as essential inputs to the simulation of bacterial population biology and ecology (Kreft et al., 1998, 2001; Ginovart et al., 2005; Prats et al. 2006). COSMIC-Rules is based on the COSMIC (Gregory et al. 2004a, 2004b, 2005; Paton et al. 2005) and RUBAM (Vlachos et al., 2005) models, which were designed to create detailed simulations that were biologically accurate and computable within practical time frames (Gregory et al., 2006). However, where COSMIC had relatively simplistic yet explicit implementations of gene

* Corresponding author.
E-mail addresses: greg@csc.liv.ac.uk (R. Gregory), v.a.saunders@ljmu.ac.uk (V.A. Saunders), jrs@liverpool.ac.uk (J.R. Saunders).

0303-2647/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2007.09.002

expression and evolution, COSMIC-Rules is rule-based with implicit implementations of gene expression that allow the compression of genetic information and facilitate more rapid computation. Accordingly, the model is designed to simulate genetic interactions in bacteria or other microorganisms within a framework that allows evolutionary processes to be observed.

The COSMIC-Rules model is thus a simplification of conditions in the real world. It incorporates three different levels, the bacterial genome, the bacterial cell and an environment inhabited by those cells. The virtual environment in COSMIC-Rules can contain multiple substances/substrates, whose relative nutrient status and/or toxicity is specified solely by the bacterial genome. Source code and movies of the simulations described here are available from: http://www.csc.liv.ac.uk/~greg/biosys/model/

2. Methods

COSMIC-Rules is built to support the modelling of evolution with mixed species of virtual bacteria incorporating available biological information to inform simulations. Each of the three levels represented in the model (namely the environment, the cell and the genome) uses the Individual-based-Model (IbM) philosophy; the cells have individual genomes with their own genes. The environment consists of multiple individual substances with heterogeneous concentrations. In this virtual world, bacteria may be subject to simple mutation events and/or other forms of genome rearrangement, for example through the acquisition of mobile genetic elements such as plasmids, transposable elements or bacteriophages (bacterial viruses). The simulation provides a potentially multi-substance environment that is populated with species of bacteria. Each bacterial cell is an individual, but genetic coding varies within a given species generating preferences for some substances and sensitivity to others.

2.1. Rules

The behaviour of the virtual bacteria, gene expression and metabolism are governed by a rule base that defines response to changing environmental conditions. Rules are used to provide the likelihood of a quantitative outcome for interactions between individual bacteria and other bacterial species, plasmids, phages and/or their environment. The rules reflect biological reality and physical laws. They are modulated by the interaction of bit strings that represent genes (see Section 2.4) and are set within specific probability values of an outcome. Potential variation occurs at every iteration in a cell line, in turn producing variability at the population level. The rules define possible interactions between a given phenotype and the prevailing environment. The genome, the cell and the environment all provide information sources for rules, but rules are not part of these elements, rules are applied from outside and are static for the lifetime of the simulation. Specific rules, tailor-made for specific functions, may be introduced for different case studies using the model.

2.2. The environment

The environment consists of multiple, individual substances with heterogeneous concentrations in a three-dimensional space, with the third dimension being only one cell diameter deep. In the simulations, organisms can be observed to move in two dimensions. This mimics the real world observations of bacterial cells in a "liquid" environment between a microscope slide and coverslip, with cells out of the plane of focus being invisible.

A substance/substrate may be treated as equivalent to another substance, or it might be unique. Each substance can be either or both a beneficial food source and a harmful or toxic/fitness-reducing agent. Moreover, it can conceivably be both beneficial and antagonistic to the same cell, in which case the cell would need to acquire resistance before utilising the substance as a food source. The substances exist in a 20 mm^2 world with toroidal edges. It is fine grained and discrete, with a 512×512 grid of floating point numbers representing the concentration of each substance (see Table 1), each (non-discrete) cell location maps to a discrete local substance concentration. As a result the overall environment appears continuous. Each substance also has an effectiveness coefficient, which changes the global concentration without destroying local information. The simulation also allows concentrations to be artificially modified at given times.

Table 1
Simulation parameters

Parameter description	Parameter	Unit/domain
Initial bacterial population	4000	Individuals
Peak number of bacteria	980,000	Individuals
Environment space	3D	Topology
Substance concentration	2D array	Floating point numbers
Floating point numbers per dimension of array	512	4 byte floats
Default world size	0.02	Meters
Simulation step size (iteration)	1	Second
Generation time (shortest)	1200	Iterations
	20	Minutes
Visualisation step size	100	Seconds
Maximum cell movement per time step	25	μm
Maximum distance for plasmid/phage transfer	12	μm
Tag size	16	Bits per tag
Default allowed bit difference	1	Bit
Mutation rate	10^{-8}	Per bit per tag per second