Modeling and evolving biochemical networks: insights into communication and computation from the biological domain

James Decraene, George G. Mitchell, Barry McMullin

Artificial Life Laboratory - Research Institute for Networks and Communications Engineering
School of Electronic Engineering - Dublin City University, Ireland
Email: james.decraene@eeng.dcu.ie

Keywords: Evolutionary computation, biochemical networks, crosstalk, adaptive control systems.

Abstract

This paper is concerned with the modeling and evolving of Cell Signaling Networks (CSNs) *in silico*. CSNs are complex biochemical networks responsible for the coordination of cellular activities. We examine the possibility to computationally evolve and simulate Artificial Cell Signaling Networks (ACSNs) by means of Evolutionary Computation techniques. From a practical point of view, realizing and evolving ACSNs may provide novel computational paradigms for a variety of application areas. For example, understanding some inherent properties of CSNs such as crosstalk may be of interest: A potential benefit of engineering crosstalking systems is that it allows the modification of a specific process according to the state of other processes in the system. This is clearly necessary in order to achieve complex control tasks. This work may also contribute to the biological understanding of the origins and evolution of real CSNs. An introduction to CSNs is first provided, in which we describe the potential applications of modeling and evolving these biochemical networks *in silico*. We then review the different classes of techniques to model CSNs, this is followed by a presentation of two alternative approaches employed to evolve CSNs within the ESIGNET project [1]. Results obtained with these methods are summarized and discussed.

1 Introduction

Cell signaling networks (CSNs) are bio-chemical systems of interacting molecules in cells. Typically, these systems take as inputs chemical signals generated within the cell or communicated from outside. These trigger a cascade of chemical reactions that result in changes of the state of the cell and/or generate some (chemical) output. CSNs can, therefore, be regarded as special purpose computers [1]. In contrast to conventional silicon-based computers, the computation in CSNs is not realized by electronic circuits, but by chemically reacting molecules in the cell. The most important molecular components of CSNs are proteins. There are many different proteins, each of which can engage in interactions with other molecules with a high degree of specificity. Their properties are often modified through interaction with other molecules. Often different CSNs are connected to one another through shared components. This is referred to as crosstalk. A potentially useful attribute of biochemical networks. Through evolution, these biochemical networks increasingly became responsible for complex adaptive control phenomena in the biological systems. These phenomena included the coordinating of cellular activities such as: to order programmed cell death (apopstosis), cell differentiation, bacterial chemotaxis etc.
In the remainder of this paper, we explore the potential applications of CSNs as engineered computational devices. We then review the different modeling techniques to specify CSNs. Following this, we describe two distinct evolutionary approaches to realize and evolve ACSNs. We finally summarize and discuss the results obtained with these evolutionary platforms.

[1] ESIGNET: Evolving Cell Signaling Networks *in silico*, an EU FP6 project, contract no. 12789, http://www.esignet.net

2 CSNs as computational devices

Realizing and evolving Artificial Cell Signaling Networks (ACSNs) may provide new computational paradigms for a variety of application areas. In this section we present some evidences found in the literature which indicate that CSNs may have natural and dedicated applications. The state of the art on using CSNs as computational devices *in silico* and *in vivo* is then presented.

2.1 Engineering crosstalk:

In this section we examine a natural phenomenon occurring in CSNs called "crosstalk" and its potential contributions to engineering. Crosstalk phenomena happen when signals from different pathways become mixed together [8]. This arises very naturally in CSNs due to the fact that the molecules from all pathways may share the same physical reaction space (the cell). Depending on the relative specificities of the reactions there is then an automatic potential for any given molecular species to contribute to signal levels in multiple pathways. In traditional communications and signal processing engineering, crosstalk is regarded as a defect: An *unintended* interaction between signals, that therefore has the potential to cause system malfunction. This can also clearly be the case of crosstalk in real CSNs, for example cells may become cancerous due to undesired

crosstalk connections [15]. However, in the specific case of CSN's, crosstalk also has additional potential functionality, which may actually be constructive:

- Even where an interfering signal is, in effect, adding uncorrelated "noise" to a functional signal, this may sometimes improve overall system behavior. This is well known in conventional control systems engineering in the form of so-called "dither". Molecular biologists indicated that noise is an inevitable by-product of inherent molecular interactions, and that in fact noise is essential for development [23].

- The crosstalk mechanism may also provide a very generic way of creating a large space of possible modifications or interactions between signaling pathways. Thus, although many cases of crosstalk may be immediately negative in their impact, crosstalk may still be a key mechanism in enabling incremental evolutionary search for more elaborate or complex cell signaling networks. For example, Genoud et al. [10] presented a number of crosstalk connections between real signaling networks occurring in plants in which these "interferences" provided a relatively *rapid* and *efficient* mechanism for optimizing non-cognitive behavior in response to various combinations of stimuli.

Both above cases of crosstalk may give new insights on the use of crosstalk in control engineering.

2.2 Computation in CSNs:

It is believed that operations similar to traditional signal processing functions exist in a number of real CSNs. An early work given by Bray [1] showed that molecules could be regarded as computational devices, these molecules could perform simple computational tasks. Examples of such computational functions are: signal acceleration, signal amplification or decision making. A review on the computational abilities of signaling networks can be found in [21]. These identified computational processes occurring in CSNs indicate that complex operational features have been designed in CSNs through natural evolution. This review highlights the computational power of real CSNs and suggests the possibility to realize and evolve ACSNs to carry out similar but *pre-specified* computational tasks.

2.3 Computing with CSNs:

In the above part, we presented different cases where computational processes were identified in real CSNs. In here, we describe the opposite approach CSNs are used to carry out pre-specified computations. Two form of scientific experimentations could be performed either *in vivo* or *in silico*.

- *In vivo* computation: Also called "Molecular computing", this approach is concerned with the realization of nano-scale computational devices using biomolecular components. So far, molecular devices such as enzyme transistors and biological logic gates have been developed [19, 4]. Nonetheless, a molecular level analog computer, in the form of a CSN may offer capabilities for high speed and small size that cannot be realized with solid state electronic technology. More critically, where it is required to interface computation with chemical interaction, a CSN may bypass difficult stages of signal transduction that would otherwise be required. This could have direct application in so-called "smart drugs" and other biomedical interventions.

- *In silico* computation: The most significant work to date regarding the use of CSNs to perform computation *in silico* has been produced by Deckard and Saura [5], in which evolutionary techniques were used to construct (simulated) biochemical networks (where reactions were represented by ODEs) capable of certain simple forms of signal-processing such as a square root function. This work also illustrated the evolution of modularity in biochemical networks, this highlights the potential of computational approaches to assist in the understanding of real phenomena.

Due to technical and financial constraints, the *in vivo* approach is still only in its infancy whereas its *in silico* counterpart has already provided significant results and insights for the understanding of real CSNs. In the next section in which we review the different techniques to represent CSNs.

3 Modeling CSNs

The literature on modeling biochemical networks is growing rapidly and the motivations behind different modeling techniques are sometimes quite distant from each other. To clarify the current context, we present a systematic overview of the different philosophies to model biochemical networks. We place a particular emphasis on three main domains which have been playing a major role in the past, namely: analytic, stochastic and algebraic approaches. For each approach, we present the key ideas and assumptions:

3.1 Analytic approaches:

These approaches mainly rely on the use of differential equations, which is currently the most widely used technique to model CSNs, see [9] for review. With differential equations, the state of CSNs are expressed in terms of molecular concentrations without inner structure. Thus the individual behavior of molecules is not considered but rather the behavior of molecular species as a whole. The use of differential equations also implies a progression of time along the x-axis, it is then possible to calculate the molecular concentrations at any given time.
When reactions are considered in a well-stirred reactor (i.e., without any consideration of space or compartmentalization) then Ordinary Differential Equations (ODEs) are commonly used. To model space, Partial Differential Equations (PDEs) are employed and involve a complementary variable denoting space. Using PDEs has a significant cost on

computation resources, an alternative is to use compartmental models which are based on ODEs but also include fluxes between compartments. This way, we may obtain a coarse representation of space and still keep the benefits of using ODEs. Because differential equations are well established in the sciences, a plethora of analytical and simulation tools exist and facilitate the modeling of biochemical networks.

3.2 Stochastic approaches:

Biochemical processes are stochastic by nature, in order to account for the uncertainties occurring during signal transduction, one may employ statistical approaches in which three principal techniques can be distinguished: Bayesian networks, Stochastic simulation algorithm and Markov chains, see [22] for review. These approaches consider the standard deviations from the average observed behavior of molecular processes. Probabilities are introduced to weight alternative behaviors that may occur. These models also take into account probability distributions and their interpretation.

Bayesian networks have recently attracted the attention of scientists for their ability to infer Signaling Networks from experimental data [17]. In Bayesian networks, the molecular species are represented as variables which are associated with probability tables. The latter indicate the different possible states (concentration level) according to other molecular species concentrations. Bayesian networks can also be used even when only steady state-data are available, in which case kinetic models are less useful [24].

Stochastic simulation algorithms (SSAs) are based on the chemical master equation and were pioneered by Gillespie [11]. SSAs are the most commonly used statistical techniques for modeling biochemical networks. In SSAs, the kinetic rate of a reaction corresponds to the probability of a reaction to occur within a time interval.

Markov chains have also been employed to represent biochemical processes, here, the state of the chain is designated by the number of molecules present. The reactions are modeled as transitions between these states. Using Markov chains, it is possible to obtain information regarding the steady-state probability distribution of a signal transduction process, however this can be carried out only if there is no feedback in the system.

As for the differential equations, statistical approach are well grounded techniques and are provided with powerful analysis tools. However, these techniques also suffer from their expensive computational cost when simulated.

3.3 Algebraic approaches:

Whereas both analytic and stochastic methods have been heavily studied for modeling biochemical networks in the literature, the algebraic approaches have not received as much attention from the System Biology community. Within the algebraic approaches, two main modeling families can be distinguished: rewriting systems and process calculi. These approaches have in common the assumption of having a finite or recursively enumerable number of atomic objects. These objects can be composed in hierarchical systems where objects are molecules and/or interactions between molecules. This approach reflects the discrete characteristics of CSNs and facilitates molecular tracing.

Term rewriting systems are a well established principle of theoretical Computer Science where molecular species are interpreted as objects represented by strings of characters (terms). Sets of term rewriting rules can be defined to describe the interactions that may occur between the objects. Information on the state of the system can be obtained from observing the state of each term. Examples of term rewriting systems are: grammar systems, P-systems [16] and classifier systems [12].

Biological processes are concurrent by nature, modeling concurrency is facilitated by process calculi where an emphasis is given on interaction modeling, communication and synchronization between concurrent computational processes (molecules). Complementary structural and chemical determinants correspond to communication channels. Chemical interactions and subsequent modifications coincide with communication and channel transmission.

Term rewriting system and process calculi provide a highly detailed description of signaling networks. However, these approaches only allow a semi-quantitative view of the system as a significant factor to be considered is the lack of an associated temporal dimension. Examples of process calculi are Petri nets [18] and π-calculus [20].

4 Evolving CSNs *in silico*

We now present two distinct and alternative evolutionary approaches which have been developed to evolve ACSNs. Both techniques are being developed within the ESIGNET project. The key difference between these two methods is that one is addressed from a top-down point of view and the second technique realizes and evolves ACSNs from a bottom-up perspective.

4.1 A top-down approach

The ESIGNET project has built a software tool called The SBMLevolver which is capable of evolving artificial biochemical networks performing pre-specified tasks. As a representation format, the systems biology standard SBML was selected due to it being a common interchange format for biochemical models. The SBMLevolver is a two-level evolutionary system constructed in a distributed architecture. The motivation for these two levels is so as to separate two distinct evolutionary components: 1) network structural evolution and 2) the kinetic parameter fitting evolution, in constructing the system in this manner it was found that kinetic parameters adapt to the mutated network structure. Furthermore a different evolutionary computation is used for the kinetic parameter fitting level. This level uses an evolutionary strategy to search and identify the optimal parameter settings The systems uses eight different mutations with respect to the differing levels (at the network and

molecular level) as follows: addition/deletion of a species, addition/deletion of a reaction, connection/removal of an existing species to/from a reaction, connection/removal of an existing species to/from a reaction, duplication of a species with all its reactions and mutation of a randomly selected kinetic parameter by addition of a Gaussian variable.

The combined use of these kinds of mutation is novel. While some previous systems have utilized some of these mutation techniques, either singly or with others, however none to our knowledge have been used with a species duplication mutation technique or applied to the kind of biochemical network evolution discussed here as far as we are aware. Fitness evaluation in the algorithm is done by integrating the ODE system resulting from an individual model. Results published elsewhere [14] have shown that the SBMLevolver is effective in the evolution of CSNs and other biochemical networks. Simple mathematical functions such as a third root computation were successfully evolved using this evolutionary system [14]. When utilized to evolve ACSNs so as to mirror real biochemical networks, it was found that implausible solutions (from a biological point of view) could be obtained. These results highlighted the requirements for the determination of better evolutionary constraints to evolve biologically plausible ACSNs. These findings would support the arguments outlined by Chu [2] where he identified a number of currently open problems regarding the methodologies employed to evolve artificial biochemical networks.

4.2 A bottom-up approach

In this section we present an alternative evolutionary approach which is also being developed in the ESIGNET project. This technique examines the emergence of biochemical networks in undirected, self-engineered and autonomous, systems. In the SBMLevolver, the evolutionary process was driven from a top-down point of view, an explicit fitness function was defined and was responsible for directing the evolutionary exploration. We introduce our bottom-up CSN evolutionary system: As CSNs occur in cells, these networks have to replicate themselves prior to the cellular division. This allows the replicated CSNs to be "distributed" to the offspring cells. The "fitness" of a cell is implicitly represented by the *survival* and *performance* of a cell in achieving self-maintenance *and* cell-level replication. Based on the above assumption, we hypothesize that CSNs may be regarded as subsets of closed (and thus self-maintaining) systems. The signal processing ability of CSNs would emerge from the closure properties of these systems.

As opposed to traditional string-rewriting systems, operations are stochastic and reflexive (no distinction made between operands and operators). The behavior of the condition (binding) properties and action events (enzymatic functions) is defined by a language specified within the MCS. This "chemical" language defines and constrains the complexity of the chemical reactions that may be modeled and simulated. These reactions result from successful molecular interactions which occur at random. A molecule may contain several condition/action rules which define the binding and enzymatic properties. We proposed a simplification of the broadcast language (BL) [6] (Learning Classifier Systems can be seen as a simplification of the BL) which is used as the MCS chemical language resulting in the MCS.b system. A detailed description is omitted in this paper, see [13] for full specification of our BL implementation.

As this system is an undirected approach, the first key step was to obtain catalytic networks that are able to self-sustain over time. In order to achieve such evolutionary robustness, it is necessary for these networks to possess mechanisms which provide some resistance to parasites that may be formed. Results obtained in our preliminary studies exhibited unexpected evolutionary dynamics which resulted in various degenerative cases: No stable cooperation between the molecular species could be observed in the evolutionary simulations [7]. This was due to the successive emergence of parasitic species which destructively overran the system and ultimately caused system extinction. This evolutionary system, suffering from a lack of robustness, was designed as a single-level Artificial Chemistry (AC) where only molecules were competing with each other. Further developments of the system included the introduction of a second level of selection [3, 7], where molecules were now contained in compartments (analogous to cells). Similarly to competing molecules, cells were subjected to artificial selection: There was a fixed number of cells, when a cell divides, another cell is picked at random and removed from the cellular population. Results obtained from this multi-level selectional system suggested that multi-level selection was an effective mean to provide resistance to parasites. Therefore this system displayed improved robustness properties. Although we succeeded to solve this stability problem and provided complementary insights into the understanding of evolutionary dynamics in minimal artificial chemistries, this highlights the fact that there is currently no theoretical framework for the study of ACs. Ongoing research investigates several scenarios (i.e., novel cellular division criteria are being explored) which would allow the emergence of regulatory feedbacks in these self-maintaining catalytic networks. This work may ultimately give rise to the emergence of *minimalist* CSNs capable of some well known engineered signal processing-like features.

5 Conclusion

We introduced some of the potential applications that modeling and evolving CSNs may provide in computation and network engineering. We reviewed the different classes of techniques to model these biochemical networks. Following this, we described two alternative evolutionary systems to evolve CSNs from the top-down (SBMLevolver) and bottom-up (MCS.b). Preliminary results obtained with these systems were briefly presented: The SMBLevolver successfully evolved ACSNs to perform a range of mathematical functions. Similarly the MCS.b evolved interesting

molecular organizations, however the evolutionary dynamics were unexpected, due to a deficit in evolutionary stability of the system. Additional developments enhanced the system robustness and allowed catalytic networks to self-sustain over time. To evolve biologically plausible ACSNs with the SMBLevolver, we proposed that developing more precise evolutionary constraints must be prioritized. Then regarding the MCS.b, we plan to extend the multi-level selectional model by introducing new cellular division criteria, which would constrain and drive the evolution of the molecular networks.

Acknowledgements

This work was funded by ESIGNET (Evolving Cell Signaling Networks in Silico), an European Integrated Project in the EU FP6 NEST Initiative (contract no. 12789). We are also grateful to fruitful discussions and comments from Ciaran Kelly and to Thorsten Lenser for his contribution on the SBMLevolver

References

[1] D. Bray. Protein molecules as computational elements in living cells. *Nature*, 376(6538):307–312, 1995.

[2] D Chu. Evolving Genetic Regulatory Networks for Systems Biology. In *Proceedings of Congress on Evolutionary Computation 2007, Singapore Sept 25-27 2007*. IEEE Explore, September 2007.

[3] M.B. Cronhjort and C. Blomberg. Cluster compartmentalization may provide resistance to parasites for catalytic networks. *Physica D: Nonlinear Phenomena*, 101(3-4):289–298, 1997.

[4] A.P. de Silva and N.D. McClenaghan. Molecular-Scale Logic Gates. *Chemistry- A European Journal*, 10(3):574–586, 2004.

[5] A. Deckard and H.M. Sauro. Preliminary studies on the in silico evolution of biochemical networks. *Chembiochem*, 5(10):1423–1431, October 2004.

[6] J. Decraene, G. G. Mitchell, and B. McMullin. Evolving artificial cell signaling networks: Perspectives and methods. In *Advances in Biologically Inspired Information Systems*, volume 69, pages 165–184. Springer, 2007.

[7] J. Decraene, G. G. Mitchell, and B. McMullin. Unexpected Evolutionary Dynamics in a String-Based Artificial Chemistry. *Artificial Life XI Proceedings*, 2008. To appear.

[8] J.E. Dumont, F. Pécasse, and C. Maenhaut. Crosstalk and specificity in signalling Are we crosstalking ourselves into general confusion? *Cellular Signalling*, 13(7):457–463, 2001.

[9] N.J. Eungdamrong and R. Iyengar. Modeling cell signaling networks. *Biol Cell*, 96(5):355–362, Jun 2004.

[10] T. Genoud and J-P. Mtraux. Crosstalk in plant cell signaling: structure and function of the genetic network. *Trends in plant science*, 4(9):503–507, December 1999.

[11] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *JPC*, 81(25):2340–2361, 1977.

[12] J.H. Holland. Exploring the evolution of complexity in signaling networks. *Complexity*, 7(2):34–45, 2002.

[13] J.Decraene. The Holland Broadcast Language. Technical Report ALL-06-01, RINCE, Dublin City University, 2006. http://elm.eeng.dcu.ie/˜alife/jd/ALL-06-01/.

[14] T. Lenser, T. Hinze, B. Ibrahim, and P. Dittrich. Towards evolutionary network reconstruction tools for systems biology. In *Proc. of EvoBIO 07*, pages 132–142, 2007.

[15] M. Mukai, T. Kusama, Y. Hamanaka, T. Koga, H. Endo, M. Tatsuta Masaharu, and M. Inoue. Cross talk between apoptosis and invasion signaling in cancer cells through caspase-3 activation. *Cancer Res*, 65(20):9121–9125, Oct 2005.

[16] G. Paun. *Membrane Computing: An Introduction*. Natural Computing Series. Springer-Verlag, Berlin, 2002.

[17] D. Peer. Bayesian network analysis of signaling networks: a primer. *Sci STKE*, 2005(281), April 2005.

[18] J.L. Peterson. *Petri Net Theory and the Modelling of Systems*. Prentice Hall, 1961.

[19] M.A. Reed and J.M. Tour. Computing With Molecules. *Computing*, 2000.

[20] A. Regev, W. Silverman, and E.Y. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Pacific Symposium on Biocomputing*, pages 459–470, 2001.

[21] H.M. Sauro. The Computational Versatility of Proteomic Signaling Networks. *Current Proteomics*, 1:67–81, 2004.

[22] T.E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling in vivo reactions. *Computational Biology and Chemistry*, 28:162–178, 2004.

[23] D. Volfson, J. Marciniak, W.J. Blake, N. Ostroff, L.S. Tsimring, and J.Hasty. Origins of extrinsic variability in eukaryotic gene expression. *Nature*, December 2005.

[24] J. Woolf, W. Prudhomme, L. Daheron, Q. Daley, and A. Lauffenburger. Bayesian analysis of signaling networks governing embryonic stem cell fate decisions. *Bioinformatics*, 21(6):741–753, March 2005.