Artificial Reaction Networks

Claire E. Gerrard, John McCall, George M. Coghill and Christopher Macleod.

***Abstract*—In this paper we present a novel method of simulating cellular intelligence, the Artificial Reaction Network (ARN). The ARN can be described as a modular S-System, with some properties in common with other Systems Biology and AI techniques, including Random Boolean Networks, Petri Nets, Artificial Biochemical Networks and Artificial Neural Networks. We validate the ARN against standard biological data, and successfully apply it to simulate cellular intelligence associated with the well-characterized cell signaling network of *Escherichia coli* chemotaxis. Finally, we explore the adaptability of the ARN, as a means to develop novel AI techniques, by successfully applying the simulated *E. coli* chemotaxis to a general optimization problem.**

I. Introduction

Natural evolution has transformed the world into a resource rich in examples of elegant solutions to complex problems. However, these solutions are often hidden in layers of biochemical detail, and are consequently little understood. Cell Signaling Networks (CSNs) are an example of one such natural "solution". They refer to the network of biochemical reactions which allow communication, response and feedback within and between cells. Many scientists have reasoned that the characteristics of cellular intelligence such as recognition, classification, response, communication, learning and self-organization [1] are the result of these complex networks [2], [3].

Significant advances in biotechnology have resulted in a surge of biochemical data, allowing hidden aspects of cell signaling to be uncovered. As understanding of cell signaling becomes further developed, its significant role in cellular intelligence is emerging. Many parallels have been drawn between CSNs, computational processing and artificial intelligence techniques. For instance, their ability to perform processing analogous to Boolean logic, negative/positive feedback loops, integration, amplification, and temporal regulation [4]. However, the fact remains that no man-made system can yet compare to the degree of sophistication inherent in these networks.

Claire. E. Gerrard and John McCall are with the IDEAS Research Institute, Robert Gordon University, Aberdeen, AB25 1HG, UK (phone: +44(0)1224-26-2476; email: c.e.gerrard@rgu.ac.uk).

Geroge M. Coghill. is with the Department of Computing, University of Aberdeen, Aberdeen, AB24 3FX, UK.

Christopher Macleod is with the School of Engineering, Robert Gordon University, Aberdeen, AB25 1HG, UK.

Artificial intelligence has progressed enormously since the birth of bio-inspired approaches (for example: genetic algorithms (GAs), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) [5]), some such approaches are inspired by biochemical networks: Artificial biochemical networks [6] and Artificial Immune Systems (AIS) [5].

In this paper we focus on exploring the mechanisms of cellular intelligence to facilitate the development of novel CSN inspired AI techniques. For this purpose a new simple representation was developed: the "Artificial Reaction Network" ARN. Rather than focus on micro-molecular detail, the ARN aims to elucidate emergent behavior within a network of chemical reactions. Its biological basis is validated using real biochemical data, including simulation of the well characterized signaling network of *E.coli* chemotaxis. Furthermore, this network is examined as a source of inspiration for development of novel AI techniques.

II. Background

Nakagaki and Yamada demonstrated that the slime mould *Physarum polycephalum* was able to solve a simple maze [7]. A maze was built from plastic films set on agar gel with four possible routes of different length between two food sources. The organism eventually formed a thick plasmodial tube via the shortest pathway between the two food sources. This behavior increased its foraging capability, conserved its energy and thus increased its chances of survival. A further study by Saigusa et al showed that, when subjected to a distinct pattern of periodic environmental changes, this organism was able to learn and change its behavior in anticipation of the next stimulus [8]. The researchers argue that the behaviors illustrated in these experiments: problem solving, recalling, and anticipating events are the result of a "primitive intelligence" that emerges from the simple low-level cellular dynamics found in CSNs.

An account of how this primitive cellular intelligence arises is provided by Bray; he describes how interconnected protein units of CSNs result in a range of sophisticated processing capabilities analogous to computational components within a circuit [4]. CSNs continuously process changing environmental stimuli via this network to generate behavior suited to current conditions. Bray refers to an instantaneous set of protein concentrations as a random access memory containing an imprint of the current environmental state. The activity is determined by kinetic factors such as binding affinities or in reaction kinetic terminology: the reaction rate, reaction order and concentration of the reacting molecules. Where conditions are highly reactive, a processing unit acts like a molecular

switch giving a binary response. Such processing can be compared to that of Boolean logic. Or, in cases of lower reactivity, a unit may provide a more gradual response comparable to fuzzy logic. These processing units are linked together in cascades of protein coupled interactions with various network topological features such as feedback loops and interconnectivity and are thus capable of forming immensely complex networks. Bray claims that such a network of logical units can perform any kind of computational processing, equivalent to a finite state-machine with the same capability as a Turing machine. Evidence concerning the logical operation of protein units can be found in a number of independent studies. Stadtman et al demonstrated that the interconversion between phosphorylated and non-phosphorylated proteins can act as a flexible computational unit [9]. Similar results were documented by Arkin and Ross who examined the computational properties of enzymatic reactions [10].

Bray highlights the similarities between CSNs and ANNs. Both are examples of networked processors, simultaneously integrating and processing signals. Where weights in a neural network are set by a learning algorithm, the strength of connections within a CSN is set by natural evolution.

CSNs are the principle machinery of cellular intelligence. They may inspire new AI techniques, not only because they allow adaptive "intelligent" behavior, but also because of their intrinsic computational and processing abilities.

III. The E. coli Chemotaxis Pathway

The chemotaxis CSN of *Escherichia coli* is well characterized [11], and as such presents an ideal pathway to explore emergent properties of cell intelligence. *E. coli* have four types of transmembrane chemoreceptor proteins called methyl accepting proteins or MCPs responsible for sensing environmental chemoeffectors and a common set of cytoplasmic signaling proteins e.g. CheA, which transmit signals by reversible phosphorylation. Where no chemoeffectors are present, *E.coli* alternates between runs and tumbles, with runs lasting approximately 1 second and tumbles for 0.1 second [11]. In the presence of chemoeffectors, tumbling frequency is reduced up concentration gradients of attractants and down gradients of chemorepellents, resulting in a biased random walk. Thus, longer duration of swims in response to higher attractant gradients result in the emergence of a high level behavior characterized by net locomotion toward more favorable conditions.

To prevent the cell from being locked in either the swim or tumbling state, the cell also has a complex adaptation response. This response increases or decreases the sensitivity of the cell, depending on current ligand occupancy, by regulating the methylization of the MCP complex, so giving the cell a primitive memory.

In the two-state model [12] the MCP receptor complex is in equilibrium between two states: swim and tumble, where chemorepellents bind to the tumble form of receptor. As methylization of the MCP complex increases the receptors shift toward the tumble form of the receptor. In this form, the receptors phosphorylate CheA molecules which then transfer phosphoryl groups to aspartate residues on CheY and CheB. Phosphorylated CheY ($CheY_p$) interacts with the flagellar motor proteins triggering clockwise motor rotation (CW) resulting in a tumbling response. As $CheY_p$ concentration increases so does the tumbling frequency. CheZ is responsible for dephosphorylation of $CheY_p$. CheB and CheR are responsible for updating the methylation record and hence the adaptation response. The adaptation response drives the CSN toward its pre-stimulus equilibrium by demethylization of the MCP complex. A comprehensive description of this network is provided by Vladimirov and Sourjik [11].

IV. Related Techniques

The exploration of cellular intelligence requires a representation which focuses on high-level behaviors that emerge from CSN system dynamics, yet still capture the processing behaviors of individual reaction units. There are numerous methods of representing chemical reactions, ranging from the meticulously detailed quantum mechanical to the highly abstracted discrete Boolean models. Gilbert et al provides an excellent overview of current popular methods [13]. In this paper we shall consider only the most relevant, that is, those which capture their networked topology.

Random Boolean Networks, introduced by Kauffman, consist of a set of logical nodes, where each node corresponds to a real world object such as a gene or protein [14]. The nodes are connected to form a circuit, where the current state of each node is calculated by performing a Boolean function on its inputs. These, although focused on network dynamics, discard most unit behavior, preferring a binary switch response rather than continuous signals, and therefore cannot capture subtle system dynamics.

The Artificial Biochemical Network (AB-net) is a highly abstracted model of a CSN, intended for robotic control. It consists of a set of nodes representing protein activity, linked by weighted connections. The output of each node is a binary square-wave signal based on the input protein activities [6].

A more recent approach is the artificial biochemical neuron (AB-neuron); currently applied to phosphorylation cycles [15]. Similarly to the AB-net, it consists of a number of nodes with weighted connections. In this model the Michalis-Menton equation provides the unit output, representing the steady-state concentration of the product. Both the AB-neuron and the AB-net are simplified representations and neither capture realistic biological behavior.

Petri Nets are used extensively in several types of information processing, including modeling CSNs [16]. They work by passing tokens representing molecules between network units. In their simplest form they have similar functionality and limitations to RBNs. However, a

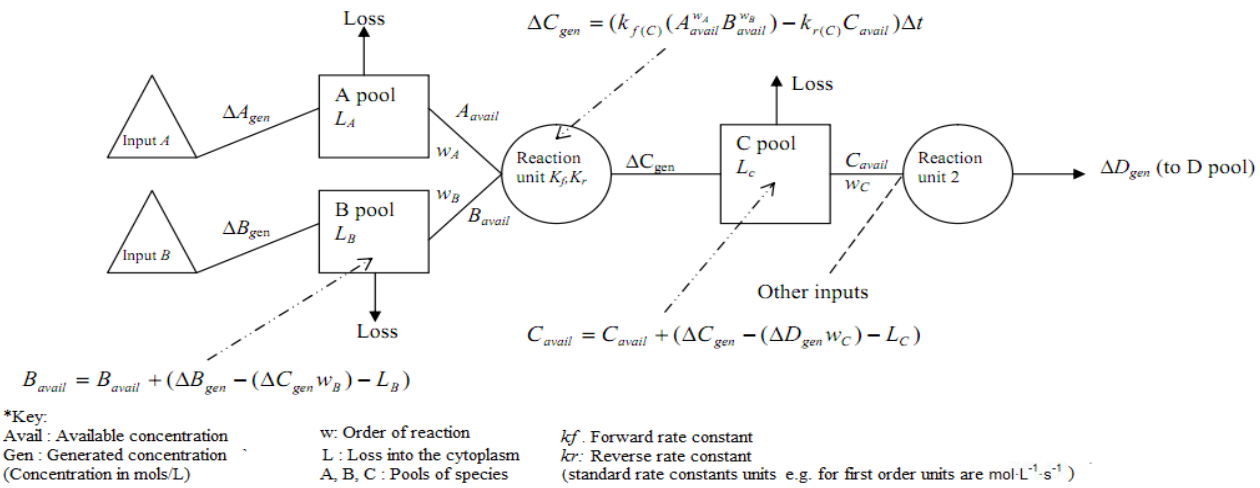


Fig. 1. The Artificial Reaction Network (ARN).

number of researchers have used them as a basis to produce more complex models.

Space precludes a complete discussion of all related models; however, it should be noted that there are several other network representations, less relevant to the problem at hand. For example, artificial immune network algorithms, and protein-protein interaction networks.

V. The Artificial Reaction Network

As explained in the previous sections, our focus is to capture the emergent cellular behavior that results from intracellular CSN processes. To achieve this, a model capable of representing sizeable networks and complex topologies, yet still maintaining biological plausibility was required. For this purpose, current methodologies were unsuitable, being either too simple or too complex, thus the authors created the ARN based on the following methods.

Developed by Savageau, S-systems are a popular representation used to model biological systems since the late 1960s [17]. They are composed of sets of ordinary differential equations (ODEs) that exploit a canonical power law representation to approximate chemical flux. Each ODE is composed of species concentration variables, raised to a power and multiplied by pseudo rate constants, as shown in Equation (1). Similarly to a traditional rate law, each ODE is equal to the difference between two conceptually distinct functions, the first term contributing to system influx, the second to decay.

To meet the previously discussed requirements, the authors combined the S-system approach with features found in RBNs and Petri Nets. By exploiting the simplified modular properties of RBNs with molecular transitions characteristic of Petri Nets, the ARN, as shown in Figure 1, represents a new, innovative, modular and expandable S-System. The ARN comprises a set of connected reaction nodes (circles), pools (squares), and inputs (triangles). Each pool represents a current species concentration (avail) measured in mols/L. Each circle represents a reaction, and calculates current flux at each time step (Δt), using Euler's approximation to the rate equation shown in Equation (1).

$$\dot{V} = k_F [A]^f [B]^g - k_R [S]^i [T]^j \quad (1)$$

Where:
$[S]^n$ = S is a species concentration, n its reaction order.
$\dot{V}$ = Current reaction rate
k_F = Forward rate constant
k_R = Reverse rate constant

Connections symbolize the flow of species into and out of reaction units and their weight (w) corresponds to reaction order. Flux (ΔA/ΔB/ΔC) as in Equation (1) and similar to S-systems, is equal to an aggregate of connected contributing (incoming) pools and connected decay (outgoing) pools raised to n powers of weighted connections and multiplied by pseudo rate constants. The pools are further subject to an optional degradation term (L), representing the natural cytoplasmic decay of species over time. This method provides each reaction with a temporal flux value, which is then used to update the current concentration values of each reaction's corresponding incoming and outgoing pools. Thus the complete set of pool concentrations at t, corresponds to the current state of the system.

The pool concept originates in Petri Nets and allows the system to account for accumulated molecular concentrations within the cytoplasm. By chaining several pools together chemical gradients and translocation through membranes can be represented; this facility is not available in standard S-systems.

Where S-systems are highly coupled sets of ODEs, the ARN is a modular approach offering finer degree of control, flexibility and adaptation. This not only supports simulation development by promoting object-orientation but is perceptually intuitive, mirroring the topology and modularization of its real-world counterpart. Thus the ARN representation is ideally suited to characterize emergent

behavior resulting from both subtle and high-level complex temporal system dynamics.

VI. RESULTS

Before the ARN could be applied to simulate cellular intelligence, its accuracy needed to be verified against known biological data and standard models. This was achieved by application of varied sets of real biochemical data to a single ARN unit. The resultant output was compared with those recorded in literature, manual calculation and by running the experiment on the Berkeley Madonna [18] programme. The outputs of these experiments confirmed its accuracy, with a minor error as expected from Euler's approximation. Figures 2 and 3 provide typical results from one such experiment. Here reaction kinetic data (rate constants, reaction order) were used to create a model of the reversible isomerisation reaction between cis and trans 1-ethyl-2-methyl cyclopropane on Berkeley Madonna and on a single ARN unit. Figure 2 shows the product output from Berkeley Madonna, and Figure 3 is that of the single ARN unit. After 2000 seconds, it can be seen that the product concentration produced by Berkeley Madonna and the single ARN are both 9.1×10^{-3} mol dm^{-3}. This result is the same as that recorded by the standard literature, thus confirming the biological plausibility of a single ARN unit.

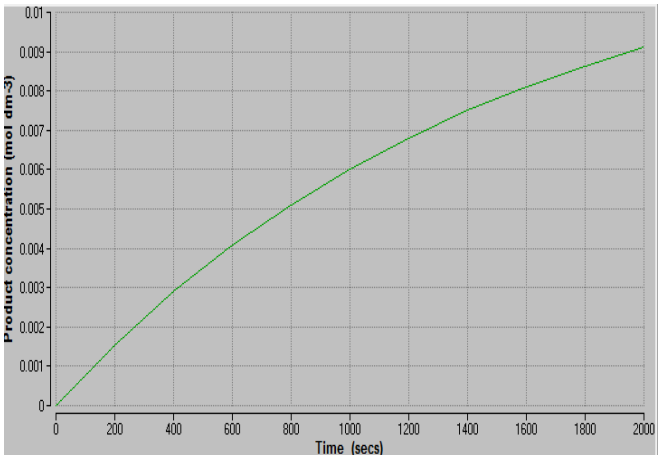


Fig. 2. The product concentration produced by Berkeley Madonna.

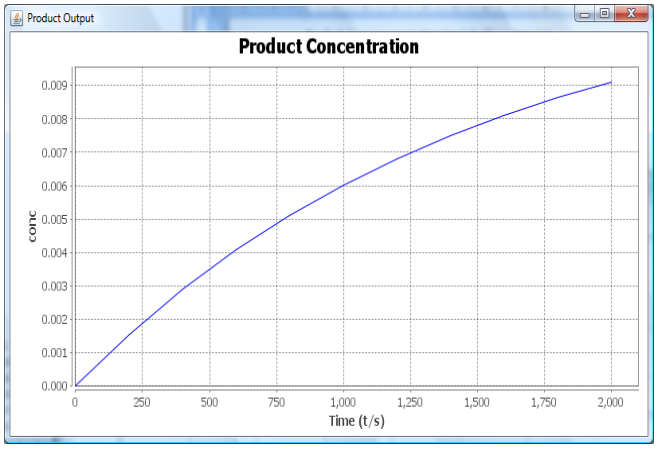


Fig. 3. The product concentration produced by the single ARN unit.

Having verified the biological plausibility of a single ARN unit, the ARN was tested as a means of capturing properties of cellular intelligence. A two state model, (refer to section3), was used as a basis to create a simulation of the chemotaxis CSN of *E. coli.* The structure of this simulation is shown in Figure 4 and is represented in the ARN format described in Figure 1 of the previous section. It is composed of a network of 10 reaction units numbered 0-9, 11 pools of intracellular signaling proteins, a single input representing the chemorepellent, and arrowed lines to show not only the connections but direction of signal flow through the network. The behavior of the simulated chemotaxis pathway in varying levels of environmental chemorepellent was setup using real biological data gathered from sources at the University of Cambridge [19], [20]. The output from this network is shown in Figures 5 and 6. Figure 5 shows the steady state concentration levels of $CheY_p$ in mols/L generated by the ARN simulation at four different continuous concentration levels of environmental chemorepellent. It can be seen from the graph that as the level of environmental chemorepellent increases so does the concentration of $CheY_p$ and therefore the tumbling frequency of the cell increases. The results are in clear agreement with published data from respected systems biology simulations [12].

To prevent the cell from being locked in either the swim or tumbling state the cell also has a complex adaptation response (refer to section 3). To ascertain the ability of the ARN to capture this behavior, the steady state concentration in mols/L of methylized MCP receptor complex obtained by the ARN simulation were examined at varying levels of continuous environmental chemorepellent.

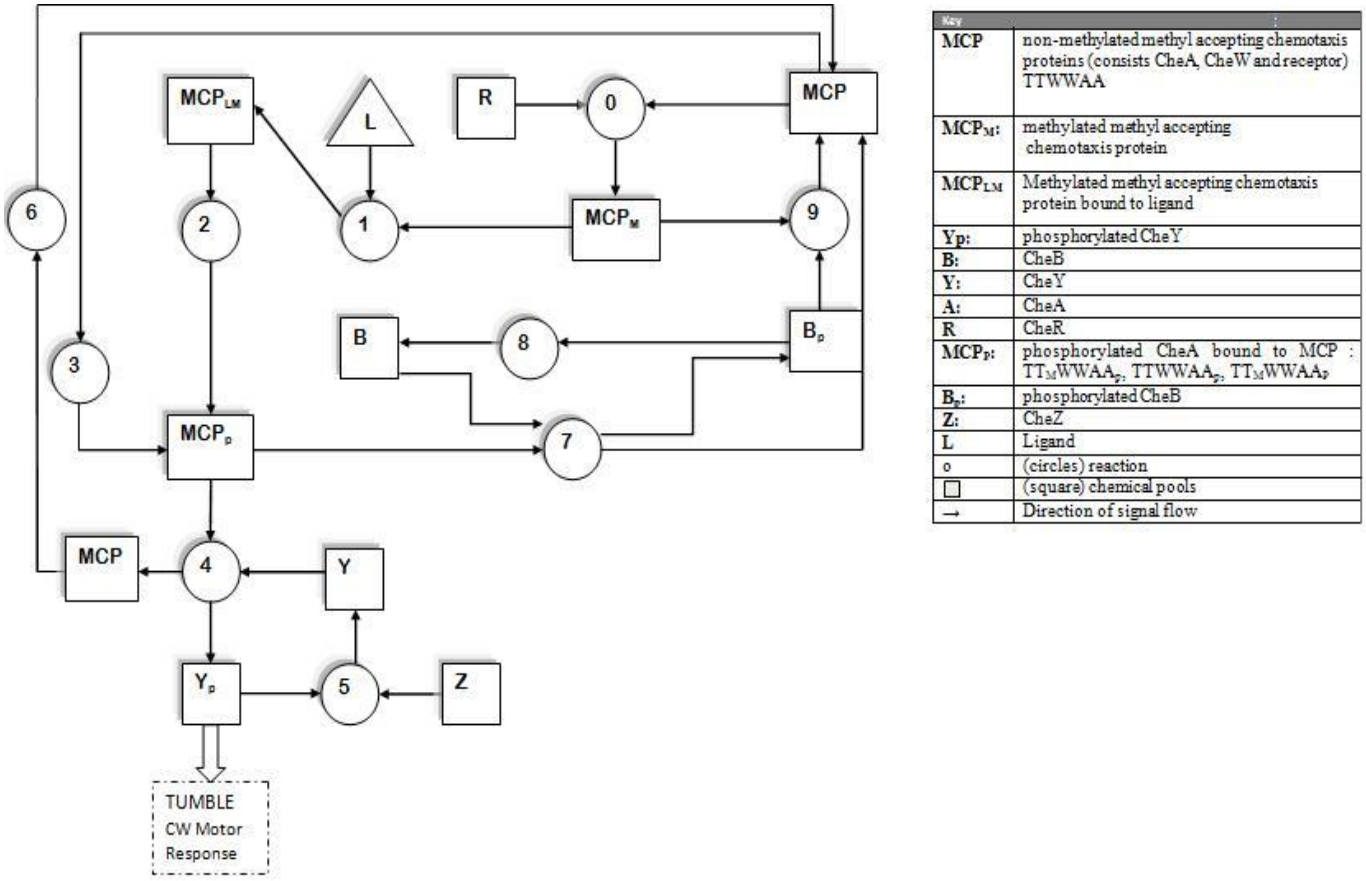
The output is displayed in Figure 6, where it can be seen that when chemorepellent concentration increases $CheY_p$ increases, and methylized MCP decreases thus driving the network back to the pre-stimulus equilibrium. Although a minor change to rate constant values were required, it can be seen that the adaptation response was attained and is in good agreement with previous work [21]

Finally to demonstrate the emergent behavior of the simulated CSN, it was decided to show the chemorepellent avoiding behavior in the context of an optimization problem. Here we observed the behavior of the simulated *E. coli* chemotaxis pathway to ascertain its ability to find a minimum chemorepellent level in an inverted bowl search space where x and y are on the horizontal plane:

$$z = \sqrt{x^2 + y^2} \tag{2}$$

Figure 7 displays the search space and an example run. The centre of the search space (solid black square) corresponds to an area of 0 chemorepellent concentration. With each progression outwards repellant concentration increases, and the outermost perimeter signifies a maximum concentration of 1×10^{-7} mols/L. The path of the simulated *E. coli* is displayed as a white line. Over 100 seconds the cell

remains in high concentration areas (above $1x10^{-9}$ mols/L) for 11 seconds and low (below $1x10^{-9}$ mols/L) for 89 seconds. These results were verified statistically over 100 run, and are in good correspondance with the reported behaviour of *E. coli* chemotaxis described in literature and using other simulation methods [21].



Key	
MCP	non-methylated methyl accepting chemotaxis proteins (consists CheA, CheW and receptor) TTWWAA
MCP_M:	methylated methyl accepting chemotaxis protein
MCP_{LM}	Methylated methyl accepting chemotaxis protein bound to ligand
Yp:	phosphorylated CheY
B:	CheB
Y:	CheY
A:	CheA
R	CheR
MCP_P:	phosphorylated CheA bound to MCP : TT_MWWAA_p, $TTWWAA_p$, TT_MWWAA_p
B_p:	phosphorylated CheB
Z:	CheZ
L	Ligand
o	(circles) reaction
□	(square) chemical pools
→	Direction of signal flow

Fig. 4. A two-state model of the chemotaxis CSN of E. coli is shown diagrammatically using the format specified in Figure 1.

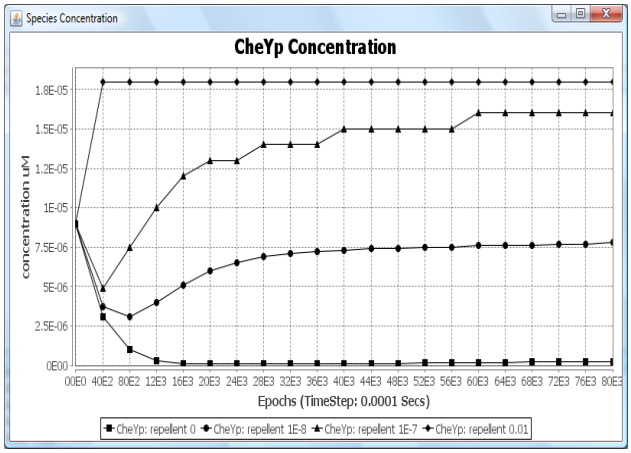


Fig. 5. The steady state concentration levels of $CheY_p$ in mols/L recorded by the ARN when subjected to varied levels of chemorepellent.

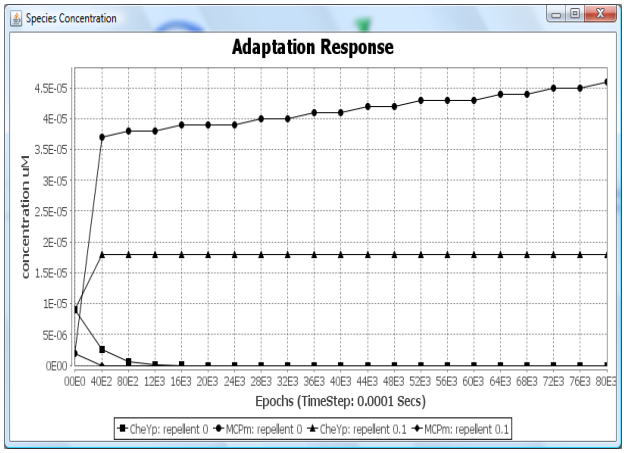


Fig. 6. The steady state concentration levels of $CheY_p$ and methylized MCP in mols/L recorded by the ARN when subjected to varied levels of chemorepellent.

Fig. 7. Minimum seeking behavior in an inverted bowl search space.

VII. CONCLUSIONS

In this paper, the ARN representation was presented as a novel method of simulating cellular intelligence. Initially, its ability to successfully represent single node reaction dynamics was shown. Its efficacy and applicability was demonstrated by creating a working model of the CSN of *E.coli* chemotaxis. This confirmed its ability to effectively simulate both the tumbling frequency regulation and adaptation response behavior of the bacteria. Furthermore, the emergent random biased walk behavior generated by the ARN was demonstrated in a general optimization problem.

The ARN approach has several advantages over other similar techniques. Its network-like structure exploits the benefits of modularization found in RBNs. It uses the molecular accounting approach of Petri Nets; however, it also incorporates the complex temporal dynamics of individual reactions found in S-Systems. The addition of pools and loss mechanisms allows more flexibility to represent intracellular compartmentalization than other techniques. The authors therefore feel that its representation is ideally suited to the characterization of emergent behaviors resulting from both subtle and high-level temporal system dynamics. Furthermore, it offers a perceptually intuitive method, as it mirrors the topology and modularization of its real-world counterpart. Aside from biological systems, this approach may also have some advantages in the simulation of other chemical systems; in particular, in the complex networks of reactions present in soil and environmental chemistry.

The modularized form of the ARN makes it particularly suitable for the application of evolutionary algorithms. The success of simulating real biological systems is generally predicated on obtaining good experimental data, which is often missing or is unreliable. Thus, the ARNs evolvability may prove useful since it promotes the identification of network parameters.

The parallels between *E. coli* chemotaxis and robotic control should be obvious. The next stage of our work involves adapting the ARN into a cellular intelligence inspired AI technique. It is intended to explore its potential as a source for development of robotic control systems and optimization techniques.

VIII. REFERENCES

[1] B. J. Ford, "Are cells Ingenious?," *The Microscope*, vol. 52, no. 3-4, pp. 135-144, 2004.

[2] K. J. Hellingwerf, "Bacterial observations: a rudimentary form of intelligence," *Trends in Microbiology*, vol. 13, no. 4, pp. 152- 158, Apr. 2005.

[3] E. Ben-Jacob, "Bacterial self-organization: co-enhancement of complexification and adaptability in a dynamic environment," *Philosophical Transactions A*, vol. 361, no. 1807, pp. 1283-1312, Jun. 2003.

[4] D. Bray, "Protein molecules as computational elements in living cells," *Nature*, vol. 376, no. 6538, pp. 307-12, Jul. 1995.

[5] J. Y. Potvin, "A review of bio-inspired algorithms for vehicle routing," in *Bio-inspired Algorithms for the Vehicle Routing Problem*, F. B. Pereira, and J. Tavares, Ed. Berlin: Springer, 2009, pp. 1-34.

[6] C. MacLeod, and N. F. Capanni, "Artificial biochemical networks: a different connectionist paradigm," *Artificial intelligence review*, vol. 33, no. 1-2 , pp. 124-135, Feb. 2010.

[7] T. Nakagaki, H. Yamada, and A. Toth, "Maze-solving by an amoeboid organism," *Nature*, vol. 407, no. 6803, pp.470-470, Sep. 2000.

[8] T. Saigusa, A. Tero, T. Nakagaki and Y. Kuramoto, "Amoebae Anticipate Periodic Events," *Physical Review Letters*, vol. 100, no. 1, pp. 1-4, Jan. 2008.

[9] E. R. Stadtman, and P.B. Chock, "Superiority of interconvertible enzyme cascades in metabolic regulation: analysis of monocyclic systems," in *Proceedings of the National Academy of Sciences USA*, vol. 74, no. 7, pp. 2761-2765, Jul. 1977.

[10] A. Arkin, and J. Ross. "Computational functions in biochemical reaction networks," *Biophysical Journal*, vol. 67, no. 2, pp. 560-578, Aug. 1994.

[11] N. Vladimirov, and V. Sourjik, "Chemotaxis: how bacteria use memory," *Journal of Biological Chemistry*, vol. 390, no. 11, pp. 1097-1104, Nov. 2009.

[12] S. Asakura and H. Honda, "Two-state model for bacterial chemoreceptor proteins. The role of multiple methylation," *Journal* of Molecular Biology, vol. 176, no. 2, pp. 349-367, Jul. 1984.

[13] D. Gilbert, H. Fub, X. Gu, R. Orton, S. Robinson, V. Vyshemirsky, M. J. Kurth, C. S. Downes, and W. Dubitzky, "Computational methodologies for modeling, analysis and simulation of signaling networks," Briefings in Bioinformatics, vol.7, no.4, pp. 339-353, Jan. 2006.

[14] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, no. 3, pp. 437-467, Mar. 1969.

[15] H. M. M. Ten Eikelder, S. P. M. Crijns, M. N. Steijart, A. M. L. Liekens, and P. A. J. Hilburs, "Computing with feedforward networks of artificial biochemical neurons," in *Natural computing Proceedings in Information and Communications Technology*, Nagoya, 2009, vol. 1, pp. 38-47.

[16] J. L. Peterson, Petri *Net Theory and the Modeling of' Systems*. Englewood Cliff, N.J: Prentice-Hall, 1981.

[17] M. A. Savageau, and E. O. Voit, "Recasting Nonlinear Differential Equations as S-Systems: A Canonical Nonlinear Form," *Mathematical Biosciences*, vol. 87, no. 1, pp. 83–115, Nov. 1987.

[18] R. Macey, G. Oster, and T. Zahnley (2011, Feb 12). Berkeley Madonna v.5, and User's Guide 2000. [Online]. Available: http://www.berkeleymadonna.com

[19] D. Bray (2011, May 20). Biochemical data (1998-2010) Bacterial chemotaxis in silico [Online] Available: http://www.pdn.cam.ac.uk/groups/comp-cell/Data.html

[20] C. J. Morton-Firth and R. B. Bourret (2011, May, 20). Experimental data in bacterial chemotaxis [Online] Available: www.pdn.cam.ac.uk/groups/comp-cell/Exp_data/exptdata.doc

[21] D. Bray, M. D. Levin, and K. Lipkow, "The chemotactic behavior of computer-based surrogate bacteria," *Current Biology*, vol. 17, no. 1, pp. 12-19, Jan. 2007.