

# Evolving Artificial Biochemical Reaction Networks: First Steps

Thomas R. Kiehl  
General Electric Research  
1 Research Circle  
Schenectady NY 12301  
kiehl@research.ge.com

Piero P. Bonissone  
General Electric Research  
1 Research Circle  
Schenectady NY 12301  
bonissone@research.ge.com

## Abstract

Evolutionary algorithms have been applied successfully to a wide variety of problems. These methods have even been applied to the biological domain. In this abstract we describe related prior work, our own system for conducting this type of *in-silico* experiment, and a discussion of our initial experiments in evolving simple functional topologies.

There have been several applications of evolutionary techniques to the domain of biochemical pathways. One of the earliest approaches evolved reaction coefficients for a given topology seeking to create a specific “dose-response” curve from the network[1]. Even with a simple fixed topology Bray, et al. demonstrated the ability to create a variety of responses including amplification, attenuation, and inversion of the input signal. While their technique was not explicitly evolutionary, it utilized a mutational operator to support a stochastic hill-climbing method used to discover a point solution. By using a true evolutionary method, one can produce a population of viable solutions rather than just a single one.

The approach used by Koza, et al.[2] sought to evolve not only the coefficients of the reactions, but also the topology of the interacting species. Ecell[3] was used to create 270 base cases. These cases reflected the behavior of four reactions from the phospholipid cycle under a variety of inputs. The fitness of an individual was based on how closely it matched the output from each of those cases. Their method, which used *Spice* to model and simulate the analogous network, would be restricted to continuous models with large concentrations. Our proposed approach can integrate stochastic models with continuous ones.

Many other applications of evolutionary algorithms to biochemical pathways have sought to find coefficients and or topologies in order to fit experimental data[4] [5]. Other unique applications in this domain include Petri nets[6], Catalytic reaction networks[7], and Combinatorial Library design[8].

Our approach is based on our simulation engine, Modesto, which is capable of executing stochastic, continuous, and hybrid simulations[9]. One significant advantage of using this type of simulator is the ability to simulate stochastic effects, which is not possible when using an analog circuit to model the system’s behavior.

Our representation is based on species and reactions. The system is given a maximal number of species that can be used to define reactions. The topology of the reaction network is represented by table 1. For instance, allele number 12 contains the forward and backward rates and the products for the reaction B + C. Alleles 6-20 represent second order reactions including dimerization reactions (such as 6, 11, etc.) First order reactions can be found in alleles 1-5, while zeroth order reaction rates, and initial concentrations can be found in alleles 21-35. This representation can be easily extended for larger numbers of species.

Reaction Network Alleles						Environmental alleles		
	A	B	C	D	E	Initial Conc	Creation Rates	Degrade Rates
n/a	1	2	3	4	5	n/a	n/a	n/a
A	6	7	8	9	10	21	22	23
B		11	12	13	14	24	25	26
C			15	16	17	27	28	29
D				18	19	30	31	32
E					20	33	34	35

Table 1: Allele locations for information encoding

Following Lenski et al.[10], we sought to evolve simple functions to be used as building blocks for more complex features. While Lenski focused on Boolean expressions derived from evolved instructions, we focused on biochemical reactions. Unlike Lenski we complemented his mutational approach with a uniform crossover mechanism.

The objective of our experiments was to validate the proposed encoding and the integration of evolutionary algorithms with our simulation engine. We were able to generate networks with varying complexity to achieve the same objective features. Initial unconstrained experiments resulted in highly complex

solutions [figure 1a]. By introducing penalties for complexity, we were able to produce a spectrum of results including the known minimal solution. Benefits of more complex networks included a regulation mechanism that maintained equilibrium at the waypoint rather than just passing through it.

We have shown that we have a mechanism to control the complexity of resultant networks. We are also able to specify more complex objectives, which will ultimately allow us to build correspondingly more complex features. As we move to higher complexity functions we want the encoding to be able to incorporate previously evolved building blocks (simpler networks).

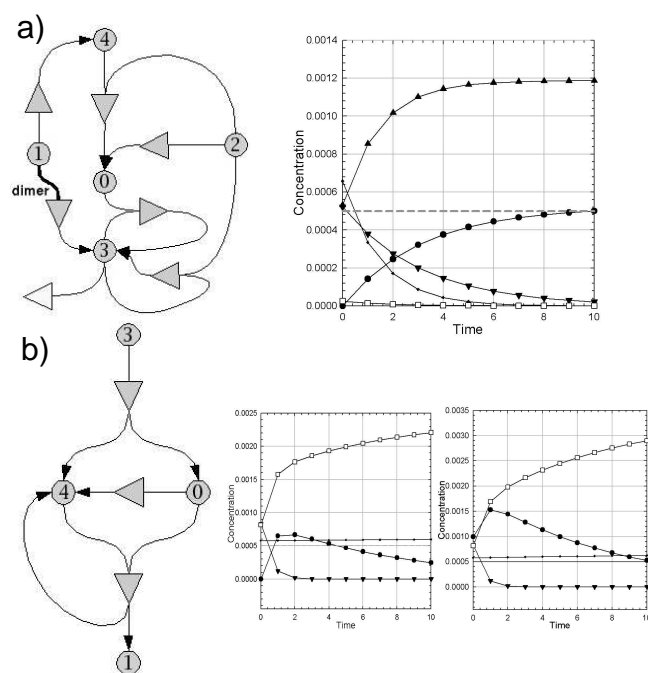


Figure 1 – a) resulting network and time course where fitness function tries to reach a specific waypoint, conc of species zero=0.0005 at  $t=10$ , from a single starting condition b) resulting network and time courses where fitness attempts to reach the same waypoint from two different initial conditions. In both charts our species of interest is species zero (●)

In our approach it is also possible to constrain a portion of the model. An experimenter may supply a-priori information about a given model such as initial concentrations, or specific reactions and rates. The system then evolves any unconstrained portion of the model. In our initial demonstration we allowed the algorithm to define all the reactions and the initial concentrations for all but one of the five allowed species.

This investment in software infrastructure paves the way for a systematic exploration of *in-silico* evolution of biochemical reaction networks. These tools, used

in an appropriate fashion, could be used to build intuition and understanding of a variety of evolutionary pressures and mechanisms. This may lend insight into understanding the environment that gives rise to the great variety of basic functions or mechanisms utilized in biological systems such as feedforward, feedback, oscillation, and bifurcation. Another interesting phenomenon to explore is the re-use of biological components and the manner in which many signals are transduced through a limited set of components. These techniques might also offer insight into the balance of efficiency and robustness of biological systems. One may even be able to use these techniques to design synthetic pathways using known molecular species.

## References

- [1] Bray D. and Lay S., *Computer Simulated Evolution of a Network of Cell-Signalling Molecules*. Biophysical Journal, 1994. **66**: p. 972-977.
- [2] John R. Koza W.M., Guido Lanza, Jessen Yu, Martin A. Keane, *Reverse Engineering of Metabolic Pathways from Observed Data Using Genetic Programming*. Pacific Symposium on Biocomputing 6, 2001.
- [3] Tomita M. and al. e., *E-CELL: software environment for whole-cell simulation*. Bioinformatics, 1999. **15**: p. 72-84.
- [4] Shinichi Kikuchi D.T., *Pathway Finding from Given Time-Courses Using Genetic Algorithms*. Genome Informatics, 2001. **12**: p. 304-305.
- [5] Michael A. Lones A.M.T., *Enzyme Genetic Programming*. Proceedings of the Congress on Evolutionary Computation 2001
- [6] Junji Kitagawa H.I., *Identifying Metabolic Pathways and Gene Regulation Networks with Evolutionary Algorithms*, in *Evolutionary Computation In Bioinformatics*, D.W.C. Gary B. Fogel, Editor. 2003, Morgan Kaufmann Publishers: San Francisco. p. 255-278.
- [7] Lohn J.D., Colombano S.P., Scargle J., Stassinopoulos D., and Haith G.L., *Evolving Catalytic Reaction Sets using Genetic Algorithms*. 1998 IEEE International Conference on Evolutionary Computation.
- [8] Weber L., *Molecular Diversity Analysis and Combinatorial Library Design*, in *Evolutionary Algorithms in Molecular Design*, D.E. Clark, Editor. 2000, Wiley-VCH: Weinheim.
- [9] Kiehl T.R., Mattheyses R.M., and Simmons M.K., *Hybrid Simulation of Cellular Behavior*. Bioinformatics, 2003(In Press).
- [10] Lenski R.E., Ofria C., Pennock R.T., and Adami C., *The evolutionary origin of complex features*. Nature, 2003. **423**: p. 139-144.