Biology: See It Again — For The First Time

Sean Luke
seanl@csl.sony.co.jp

Shugo Hamahashi
shugo@csl.sony.co.jp

Koji Kyoda
kyoda@csl.sony.co.jp

Hiroki Ueda
ueda@csl.sony.co.jp

Sony Computer Science Laboratory
3-14-13 Higashi-Gotanda, Shinagawa-ku
Tokyo 141, Japan

Computer science owes a huge debt to biological systems. The field itself came about largely as an attempt to understand and replicate the function and abilities of the brain, a complex biological creation. From this early lineage has sprung many subfields derived largely from biological metaphors: computer vision, neural networks, evolutionary computation, robotics, multi-agent studies, and much of artificial intelligence. In some areas, the computer has bested its biological counterparts in efficiency and simplicity. But for many domains, even after decades of hard work, the biological "real thing" is still superior to the artificial algorithms inspired by it.

Much of our failure to fully grasp (much less improve on) biological systems has been their sheer size and complexity. Gathering the data necessary to understand such systems *on a global level* demands tedious decades of experiment, often requiring technology that has only recently come into existence. Even if we had the data, visualizing and modelling it has historically required computing power beyond our means. Because our comprehension of these systems has been limited, many of computer science's bio-inspired fields have had to guess which salient elements of biological systems to abstract, with mixed results.

For example, in the field of applied neural networks the rich complexities of neurons, global chemical interactions, and the brain-body interface have been mostly abstracted out, leaving behind simple equations describing the flow of discrete information between nodes in graphs. The result produced interesting and useful mechanisms for pattern recognition and associative memory. But (excepting cortical map development) such simple models have not given us as much understanding of the brain or of cognition as we'd like to admit.

They've done *what?*

While computer science has by necessity been simplifying its inspirations from biology and other fields, biologists have been quietly catching up behind the scenes. It may soon be possible to model entire neurosystems, gene regulation mechanisms and evolutionary processes, even whole organisms inside a computer. These advances are due to a huge influx of data in recent years, coupled with new sophisticated computer technology.

The advances have been remarkable. Consider *Caenorhabditis elegans*, a microscopic earthworm and a popular test organism. Unbeknownst to much of the AI community, biologists a decade ago mapped out the *entire C. elegans brain*, neuron by neuron, synapse by synapse. Scientists have also identified many of *C. elegans*' neural subprocesses, worked out the entire cell lineage, sequenced much of the genome, and determined the spatial position and movement of all of *C. elegans*' embryonic cells. Biologists may soon be able to model the entire *C. elegans* brain, cell structure and development, significant chemical pathways, and physical dynamics. In many respects, this animal may become the first multicellular organism to be realistically simulated in its entirety.

This modelling can do more than just help biologists comprehend complex data and make useful experimental predictions. These sophisticated models bring new information and previously infeasible approaches to computer science fields whose early biological inspirations had been based on sparse understanding. For cognitive robotics, for example, a complete understanding of the *C. elegans* nervous and musculature system can demonstrate how little wiring is actually required (under 400 neurons!) to develop highly successful autonomous agents with simple learning and a host of complex foraging, avoidance, and mating behaviors.

Our work to date has been in two primary areas: in modelling and visualizing the cellular and neural structure of *C. elegans*, and in modelling, visualizing, and drawing evolutionary computation inspiration from the complex developmental gene regulation systems in *Drosophila melanogaster*.

C. elegans

The rich collection of available cellular and neural data on *C. elegans* makes this organism a natural subject for our research. We have begun by designing a Java-based visualization system which displays much of the non-genetic data available on *C. elegans* in a coherent manner (see Figure 1). This includes cell position and movement through time, cell lineage, cell type and expression, and neural and synaptic relationships. This allows us to see, for example, the development and movement of cells that would later become the skin, muscles, and neurons in the organism.

But visualization is only the beginning. The available spatial position data is sparse: as such we have developed a kinematic model which moves cells to proper positions, enabling us to better predict where certain cells will be at a given time. We have also begun modelling the neural system of the animal; for example its *thermotaxis*

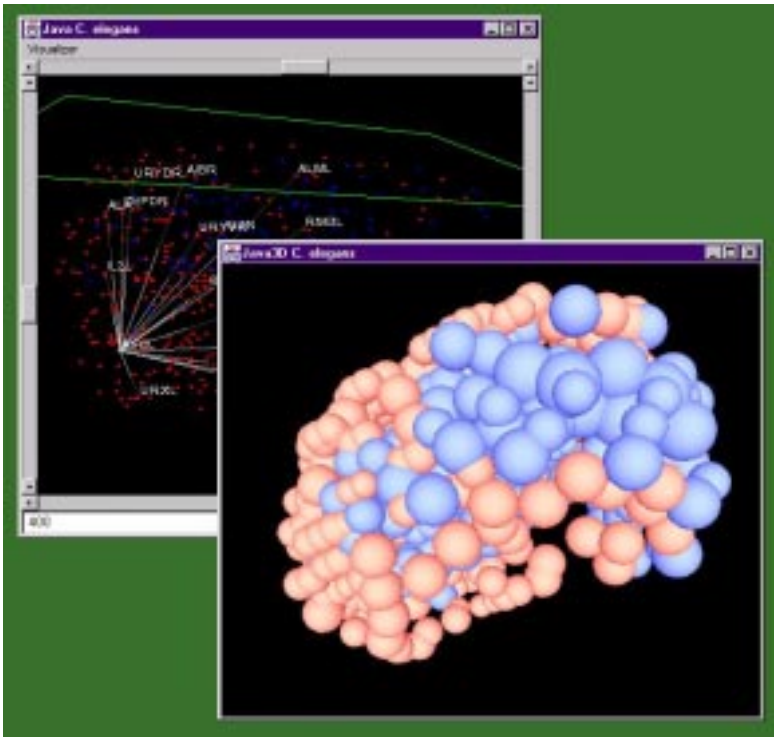

Figure 1: The Java *C. elegans* visualization system, showing volume and neural interconnection data.

circuit, which learns the temperatures at which food has been found in the past. Our immediate goal is to model the entire circuitry of the C. elegans brain, sensory organs, and musculature.

Drosophila melanogaster

Drosophila melanogaster (the housefly) is popular for genetic experimentation because many of its high-level mutations are so visually obvious. Like *C. elegans*, *Drosophila* too has a large amount of online data.

We have for some time been designing predictive models of the high-level genes involved in various aspects of *Drosophila* development. One such project describes the interaction of regulatory genes in the *Drosophila* zygote. Like humans, flies develop different body parts through *gene regulation*, a process whereby high-level genes produce proteins which control the production of lower-level genes responsible

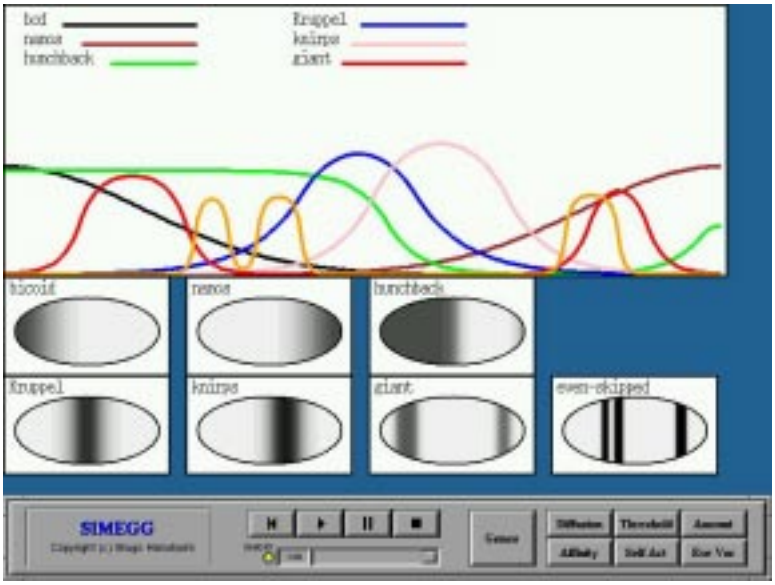

Figure 2: Simegg, the *Drosophila* zygote gene-regulation model.

for the development of specific organs and tissues. These high-level genes also regulate each other, resulting in a complex graph of interrelationships. This, plus the initial spatial distribution of specific regulatory protein concentrations in the zygote, results in a complex mosaic of protein concentrations throughout the embryo, which then give rise to different parts of the organism.

Our model successfully describes the concentration of regulatory proteins in the *Drosophila* zygote. We can mutate genes or regulatory interactions, predict the new concentration mosaics, and then compare those predictions with experimental data (Figure 2). We have gone on to create models of *Drosophila* leg growth and eye disc formation based on gene regulation and cell-to-cell chemical interaction (Figures 3 and 4).

How We Benefit

We think that a lot of ideas coming out of computer biology stand to benefit AI a great deal. Certainly working out the global mechanism behind the *C. elegans* brain can lead to better insights into neural cognition, and modelling insect eye development and circuitry might lend a hand to vision-based reactive robotics. We have investigated these and other possible avenues drawn from biology coming into its own.

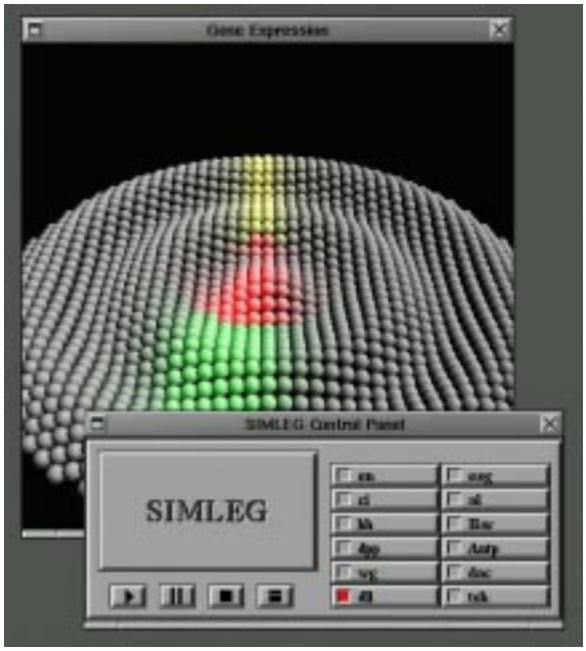

Figure 3: Simleg, the *Drosophila* leg-growth simulator.

One example: gene regulation has given us new insights into better evolutionary computation techniques. The Genetic Algorithm's (GA) approach to encoding, selection, recombination, and mutation is drawn from aging Mendelian genetics, parts of which have worked well and parts of which have not. In particular, the utility of GA recombination is now under attack (see for example [Tate and Smith 1993], [Hinterding, Gielewski and Peachey 1995], [Angeline 1997]). In real biology, recombination appears to result in mostly smooth, diverse gradations; on the other hand, in many GA domains (Genetic Programming in particular), recombination is often little more than randomization [Luke and Spector 1998]. Among other reasons, we suspect much of this may be due to the huge gap between biological and artificial GA encodings.

The approach we are now studying is directly inspired from gene regulation maps, and we are using it to evolve grammars of arbitrary symbols, neural networks, state machines, and especially Genetic Programming-style algorithms. As it turns out, gene regulation has a surprisingly one-to-one correspondence with such evolutionary domains. Our early results so far have been promising. In evolving for the Tomita language set (a benchmark for language induction), we have achieved results which compare well with other unbiased EC approaches in the literature. Using such an approach also lets us transfer many exciting and possibly useful theories from genet-

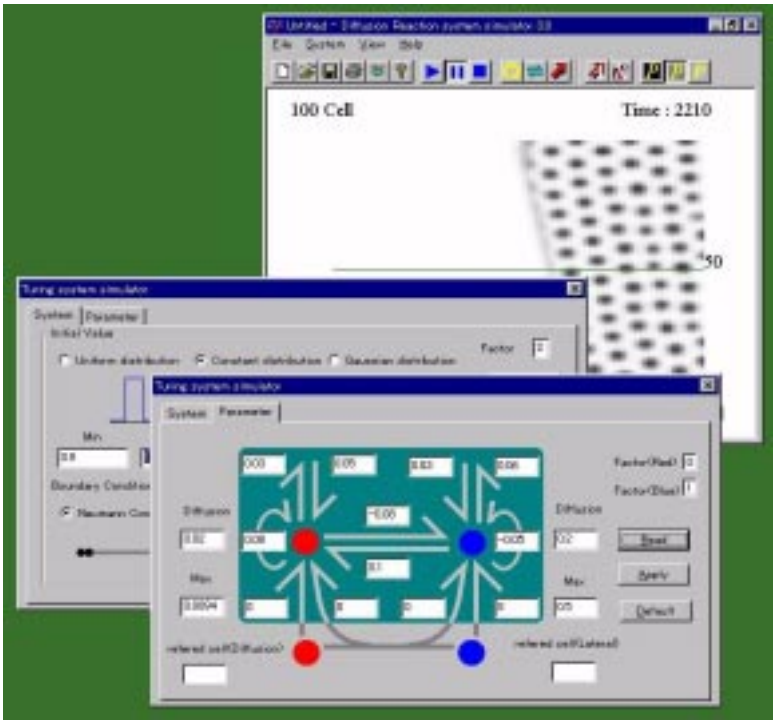

Figure 4: the gene-diffusion-reaction simulation system, modelling *Drosophila* eye-disc formation.

ics, including operon and gene family creation, transposons and gene migration, and supergene clusters.

Given how much of AI and computer science is inspired by biological metaphors, we have no doubt that biology can in similar ways help reinvigorate many other AI subfields. And modelling and analysis promise to soon make possible many things which have long been pipe dreams of autonomous robotics, artificial life, and cognitive science. It is astonishing how little AI researchers have realized how far computer-aided biology has come in the last decade. It's high time we revisited our biological roots.

Angeline, P. 1997. Subtree Crossover: Building Block Engine or Macromutation? *In Genetic Programming 1997: Proceedings of the Second Annual Conference (GP97)*. J. Koza et al, eds. San Francisco: Morgan Kaufmann. 240–248.

Hinterding, R., H. Gielewski, and T.C. Peachey. 1995. The Nature of Mutation in Genetic Algorithms. In *Proceedings of the Sixth International Conference on Genetic Algorithms*, edited by L.J. Eshelman. Morgan Kaufmann. pp. 65–72.

Luke, S. and L. Spector. 1998. A Revised Comparison of Crossover and Mutation in Genetic Program-

ming. *In Genetic Programming 1998: Proceedings of the Second Annual Conference (GP98)*. J. Koza et al, eds. San Francisco: Morgan Kaufmann. 208–213.

Tate, D.M., and A.E. Smith. 1993. Expected Allele Coverage and the Role of Mutation in Genetic Algorithms. In *Proceedings of the Fifth International Conference on Genetic Algorithms*, edited by S. Forrest. Morgan Kaufmann. pp. 31–37.