

## **Phenotypic Variability in Canalized Developmental Systems**

Sean T. Psujek<sup>1</sup> and Randall D. Beer<sup>1,2</sup>

<sup>1</sup>Department of Biology

<sup>2</sup>Department of Electrical Engineering and Computer Science  
Case Western Reserve University, Cleveland, Ohio, USA 44106  
stp4@case.edu, rxb9@case.edu

### **Abstract**

The evolution of nervous systems has been relatively conservative despite a large diversity of behavior. Novel behaviors appear to be produced during evolution by changing synaptic connectivity (Edwards and Palka 1991, Shaw and Meinertzhagen 1986, and Nishikawa et al. 1992). Due to a process of canalization, in which the development of organisms becomes buffered against genetic and environmental changes, phenotypic variation is reduced (Waddington 1942). Therefore, we explore phenotypic variability in simulations of neural development in which the developmental pathways have become canalized. Variation is determined under point mutations and gene knockouts. Although a genetic regulatory network utilizing activation and controlling a single process of neural development is evolvable and can become canalized, large genetic perturbations do not result in increased phenotypic variability as expected from related work (Bergman and Siegal 2003).

### **Introduction**

Over evolutionary time, the behavior of organisms has varied considerably despite the relatively conservative evolution of nervous systems (Nishikawa 1997). For example, insect nervous systems have a similar cellular composition yet the class shows remarkable diversity in behavior (Edwards and Palka 1991). Instead of cell type alterations, connectivity changes in nervous systems appear to play a large role in generating new behaviors in invertebrates (Edwards and Palka 1991, Shaw and Meinertzhagen 1986) as well as vertebrates (Nishikawa et al. 1992).

Connectivity changes occur as modifications in neural development; yet, developmental systems are canalized, that is, they are buffered against genetic and environmental perturbations and show decreased phenotypic variation (Waddington 1942). Phenotypic variability can be regained through large genetic and environmental perturbations which decanalize the developmental system (Rutherford and Lindquist 1998, Bergman and Siegal 2003). Do the synaptic changes that lead to novel behaviors occur as a result of small scale genetic variation (e.g. point mutations) or are they due to large genetic perturbations such as gene knockouts?

To answer our motivating question, we seek to create a model system that demonstrates canalization as well as a release from this buffering. Recently, another simulation study has demonstrated decanalization in genetic regulatory networks (GRN) through gene knockouts (Bergman and Siegal 2003). The phenotype in their study was the pattern of gene expression at equilibrium. We use the connectivity matrix of the neural network that results from the pattern of gene expression as our measure of phenotype. We aspire to have the behavior of an artificial organism as the phenotype upon which selection will be based.

In this paper, we introduce our model of neural development and demonstrate that it can reproduce the connectivity pattern of biological neural networks. We also analyze the evolvability of the model and illustrate canalization in evolved genetic regulatory networks. Using canalized and non-canalized developmental systems, we explore phenotypic variation produced under two genetic perturbations. Using this model, we hope to elucidate the mechanisms in developmental systems underlying phenotypic variability.

### **Target Identification Model of Neural Development**

There are numerous processes (e.g. cell migration, cell death, axon pathfinding, target cell identification, synaptogenesis) involved in neural development. Even though these processes all work together to create the final pattern of connectivity, neurites must make a final selection for synaptogenesis among nearby cells (Spencer et al. 2000). Therefore, we focused on the process of target cell identification to generate neural circuits.

A generalized interpretation of Sperry's chemoaffinity hypothesis is used to determine neural specificity (Sperry 1963). In our model, cell surface proteins identify the pre-synaptic and post-synaptic cell and determine the type of connection. Neurons will make synaptic connections only if the surface proteins of the cells are complementary. The post-synaptic protein determines whether the connection is inhibitory or excitatory. Dellaert and Beer (1996) and Fleischer (1995) used a similar mechanism to identify target neurons although both models incorporated the

guidance of axon growth cones as an additional developmental process. In our model, neurites are not guided and are assumed to have spread throughout the two-dimensional space in which the nervous system is grown.

Using only one process of neural development allows us to add additional complexity to the model in the direction of other neural developmental processes, regulatory mechanisms, or genetic operators (e.g. gene duplication/deletion) based on our findings.

Our interpretation of Sperry's ideas is supported by experiments in cultured neurons from the snail, *Lymnaea stagnalis* (Syed et al. 1990, Magoski and Bulloch 1998). After removing the neurons that form the respiratory central pattern generator (CPG) from the snail and culturing the cells in a brain-conditioned medium, the neurons extended neurites in all directions and formed the appropriate synaptic connections to recreate the CPG.

To simulate this process, our model of neural development proceeded through four steps. In the first step, cells were placed by the user at specific positions in a planar environment. These positions did not change during development. Following the placement of cells, a growth period occurred in which a regulatory gene network involving intra- and extra-cellular signaling proteins controlled the expression of synapse-determining surface proteins. At the end of the growth period (the length of which is arbitrarily determined by the experimenter), a list was produced of the surface proteins expressed by each cell. Finally, this list was converted by a function into the connectivity matrix for the network. The connectivity matrix indicated which neurons were connected and whether the connections were inhibitory or excitatory.

Figure 1 shows the final step of development in the formation of three synaptic connections from the pattern of expressed surface proteins. After the growth period, each cell has two surface proteins expressed represented in the figure by squares, triangles, and circles. Similar proteins have the same shape. The filled shapes are pre-synaptic proteins and open shapes are post-synaptic. A neural connection is formed between cells in which a complementary pair is expressed. The function ignores complementary surface proteins expressed in the same cell (i.e. no self-connections are formed).

Surface proteins encoded in the genome are expressed by intra- and extra-cellular signaling proteins during the growth period of development. At each time step during growth, the developmental algorithm checks each cell's intra- and extra-cellular environments for signaling proteins causing transcription events. Expressed surface proteins are then bound in the cell's membrane and do not degrade over the remainder of the growth period. If new signaling factors are expressed, they go into a new cell environment list which replaces the old one after each time step. In effect, signaling proteins act for only one time step before being degraded.

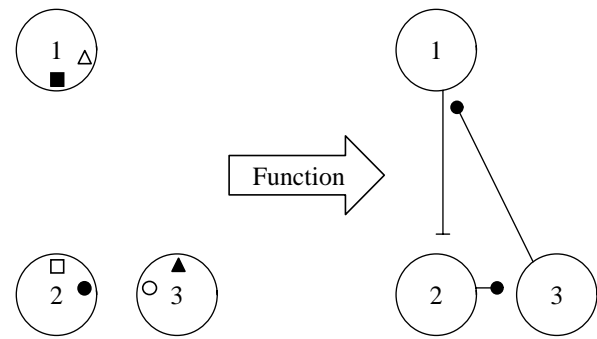


Fig. 1. Expressed surface proteins are used by a function to create the synaptic connections. Each shape in the left group of cells represents a unique surface protein. Complementary protein pairs are designated by filled and open shapes. In the right group of cells, excitatory connections are indicated by short bars, and inhibitory connections are indicated by filled circles.

Figure 2 shows a gene template of our model which encodes the various proteins and allows for gene regulation. We used two bases represented as '0' or '1', instead of the four used in DNA (adenine, guanine, cytosine, thymine), to create the functional unit of a gene. Alleles, therefore, are genes that express different proteins. There are four regions of an individual gene: activation, response range, function, and protein identifier.

Gene regions:

Activation	Response Range	Function	Protein Identifier
xxxx	xxxx	xxxxx	xxxxxx

Fig. 2. Gene template. Each 'x' represents the location of a single base. There are only two bases in our model, '1' and '0'.

The activation and response regions act as a gene's promoter by determining if a protein causes expression of a gene. An intra-cellular signaling protein need only match the activation region of a gene to cause its expression. An extra-cellular protein must not only match the activation region but must also have been expressed close enough to another cell for a particular gene to respond. This mimics a concentration gradient. The response region encodes distance and is used to determine if the distance of the origin of the extra-cellular factor is too far away for the gene to respond. This encoded distance is mutable, allowing a gene, over evolution, to respond to a chemical further away or to prevent expression if the chemical is released too close.

The functional region encodes the protein's role in development. Proteins in the model can act as signaling chemicals within and between cells or as cell surface proteins that determine synaptic connections. Intra- and extra-cellular signaling proteins result in transcription of genes. Cell surface proteins could be of three forms: a pre-synaptic, an inhibitory post-synaptic, or an excitatory post-synaptic chemical tag. Mutations in this region can result

in the function of a protein being altered, for example, turning a pre-synaptic tag into an extra-cellular signaling protein, but mutations will never result in a functionless protein.

Last in the gene sequence is the protein identifier region. The bit sequence here uniquely describes the protein expressed; there was no translation component to this model. Converting this bit sequence from base 2 to base 10 gave the protein's numerical identifier used in the examples below.

Mutations occurring in the two bits at the end of the functional and protein identifier regions do not affect the respective region's function/property and thereby add stability to the developmental system.

Transcriptional regulation of the genes is accomplished by the signaling proteins. In Figure 3, a signaling protein indicated by an oval binds to an equivalent sequence on the activation region of *Gene 1* resulting in the expression of another signaling protein, '0001'. It, in turn, activates *Gene 2* to produce a cell surface protein, '0110'.

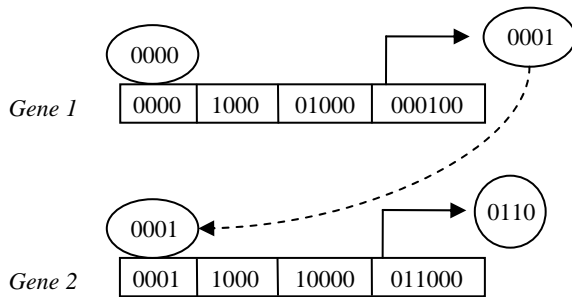


Fig. 3. Transcriptional regulation. Signaling proteins (ovals) act as transcription factors by binding to the activation region of the gene initiating the production of other signaling proteins and surface proteins (circle). Solid arrows indicate transcription of a gene's protein coding region.

Figure 4 shows a genome that creates a three-connection neural network in three cells. It will be used to demonstrate how the developmental process results in a neural network. The binding of signaling proteins to activation regions is shown by shading the protein and the bound region. '0000' is the symmetry-breaking protein introduced during the first time step. It results in transcription of genes 1, 2, and 3. The last column indicates the function and identifier of the encoded protein. For example, the bit sequence in the functional region of *Gene 2*, '1100', (minus the last two bits) denotes that the protein is an inhibitory post-synaptic tag. The protein identifier, '0010', converted to base 10 is '2'. For ease of reading, we use the English description of the protein instead of its binary sequence.

The expression pattern of this genome in three neurons is shown in Figure 5. A protein was introduced only into *Neuron 1* at the beginning of the developmental period to break the symmetry between the developing cells. The arrows indicate expression of genes with the resultant proteins at the head of the arrow. Extra1 and Extra2 are

proteins that can leave the cell and induce expression events in nearby cells (they do not result in gene expression in the cell from which they were expressed). Note that Extra1 does not result in the transcription of Post0\_Exc, Pre1, and Extra2 in *Neuron 3* due to the response region of those genes. Extra1 was expressed too far away for the genes corresponding to those proteins in *Neuron 3* to respond to the extra-cellular signaling protein. The neuronal circuit for this expression pattern is shown in Figure 1 on the right.

	Act.	Resp.	Fctn.	Id.	
1	0000	1000	10000	000000	Pre(-Synaptic)0
2	0000	1000	11000	001000	Post(-Synaptic)2_Inh
3	0000	1000	01000	000100	Extra(-cellular)1
4	0001	1000	10000	000100	Pre(-Synaptic)1
5	0001	1000	10100	000000	Post(-Synaptic)0_Exc
6	0001	1000	01000	001000	Extra(-cellular)2
7	0010	0110	10000	001000	Pre(-Synaptic)2
8	0010	0110	11000	000100	Post(-Synaptic)1_Inh

Fig. 4: Hand-designed genome for a three-connection circuit in three neurons. Shaded regions represent the signaling proteins and the activation regions to which they bind (ignoring the last two bits of the protein identifier region). The darkest transcription regions are those which are bound by the symmetry breaking protein (0000) in the first stage of development. \_Inh represents an inhibitory protein, and \_Exc represents an excitatory protein.

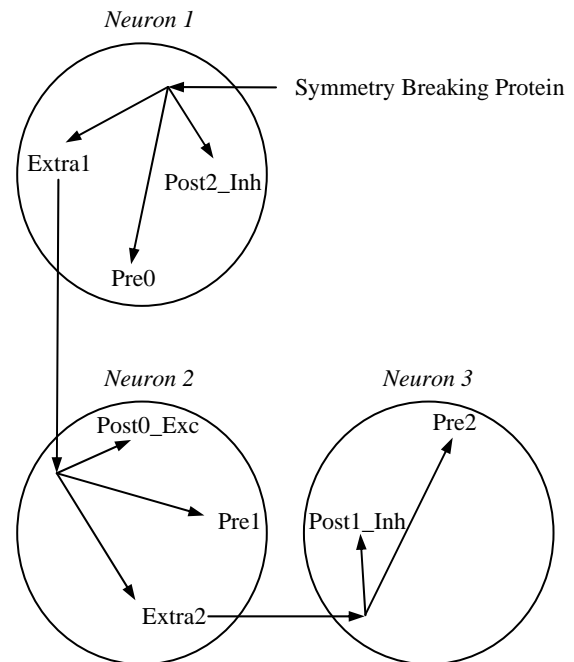


Fig. 5. Proteins expressed during development to produce the three-connection network with three neurons. Arrows indicate the cascade of expression events. Distances between cells shown above are representative of the differences in cell spacing used in development.

To demonstrate that this model is capable of reproducing the connectivity of biological neural networks,

we sought to replicate the three-interneuron *Lymnaea* respiratory CPG (Syed et al. 1990) with an additional connection from interneuron RPeD1 to a fourth interneuron, VD2/3 (Magoski and Bulloch 1998). This target network was reproduced using 15 genes. The result of development from this new genome is shown in Figure 6.<sup>1</sup>

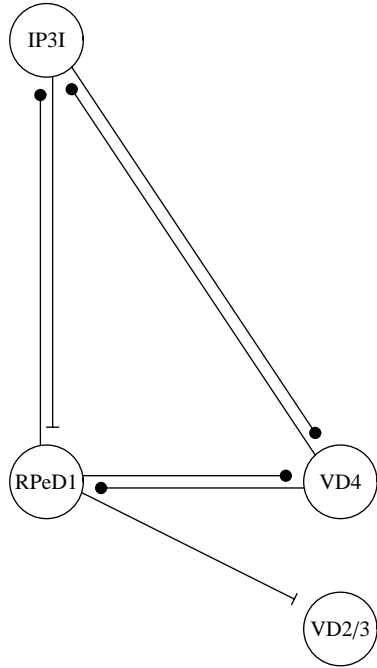


Fig. 6. The circuit architecture and cell spacing for the *Lymnaea* respiratory network. Neuron IP3I controls expiration and VD4 controls inspiration by opening and closing, respectively, the snail's pneumostome. Excitatory connections are indicated by short bars. Inhibitory connections are indicated by filled circles.

### Characterization of Single-bit Mutations

We explored the effects on phenotype of single-bit mutations in the genome that produced three-connections. For our genome of eight genes of nineteen bits each, we performed all possible one hundred fifty-two single-bit mutations and found nineteen unique architectures resulted. The majority of the mutations (fifty-nine) resulting in the original circuit. Only ten mutations resulted in no synaptic connections being formed. Ten architectures occurred only once from a point mutation. Figure 7 shows the original three-connection circuit and all of the unique architectures resulting from the point mutations.

Mutations had the effect of removing or adding synaptic connections or performing a combination of both resulting in high variability from this genome. This analysis

provides some evidence for the evolvability of the system with the caveat that it was based on a hand-designed genome. It also indicates that small scale genetic variation, i.e. single-bit mutations, can produce large phenotypic variation, the number of unique networks produced.

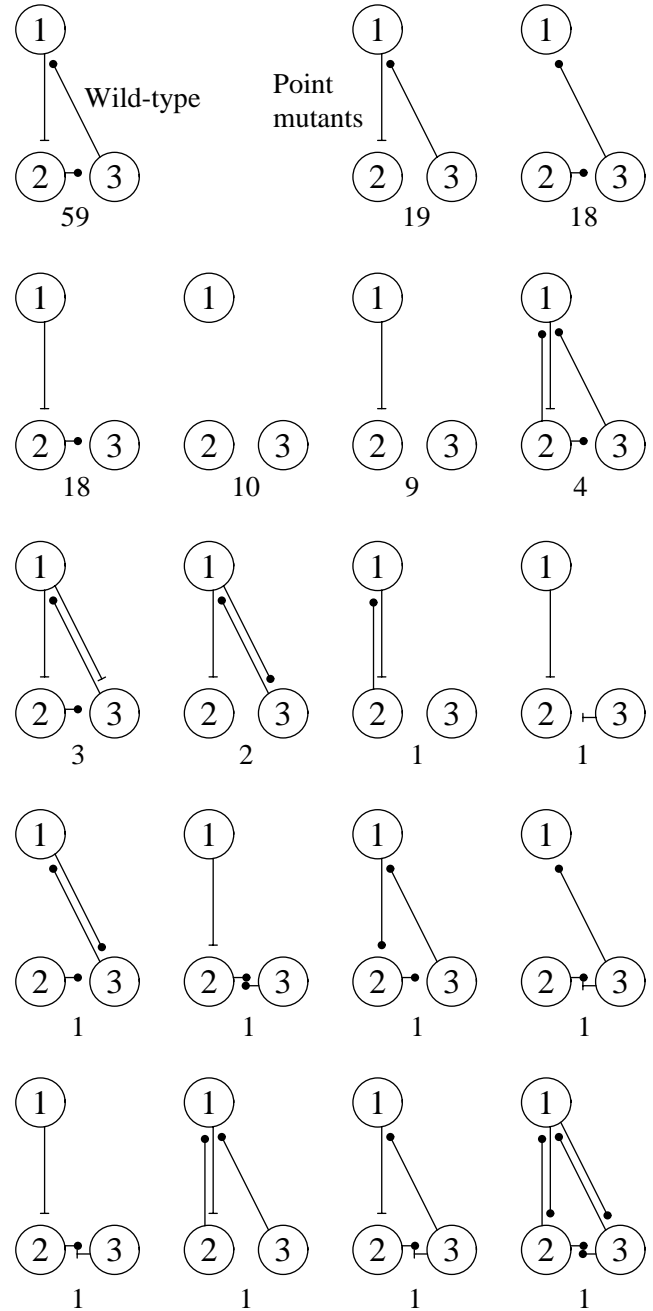


Fig. 7. The circuit architecture and cell spacing for a three-connection network and all unique architectures produced through all possible single-bit mutations. The number of mutations resulting in the network is shown below each circuit.

<sup>1</sup> RPeD1 to IP3I is actually a mixed inhibitory-excitatory connection but whose action is mainly inhibitory.

## Canalization of Development after Evolution

We performed fifty genetic algorithm (GA) runs from random initial genomes. As the hand-designed genome was composed of eight genes, we decided to use ten genes for the GA runs to allow for more possibilities of cell-cell signaling and expression of surface proteins.

We used a mutation rate that averaged one mutation per genome. No crossover or an elitist fraction was implemented. Individuals were haploid and reproduction was asexual.

As a measure of fitness, we decided to use the ratio of actual to maximum number of connections. This fitness function was chosen for its simplicity during the initial testing of evolving neural networks using the developmental model. Ignoring self-connections the maximum number of connections in a given  $n$ -neuron network is  $n^2-n$ . The ratio was entered into a parabolic function that returns a value from zero to one with a maximum value of one at a ratio of  $\frac{1}{2}$ . Selection was fitness proportionate.

Forty-one of the GA runs were able to reach the maximum fitness in three hundred generations. The other nine runs never produced synaptic connections. In those runs, the random initial genomes and subsequent mutations never resulted in the minimum expression of a pre-synaptic, a post-synaptic, and an extra-cellular signaling protein required to form at least one synaptic connection.

### Phenotypic Variation of Canalized Runs

If there is an implicit selective pressure for developmental robustness, it would be strongly evident in the sub-population of maximum fitness individuals; otherwise, the explicit selective pressure from our fitness measure would be stronger. Therefore, we calculated the average number of neutral mutations (the number of mutations that leave the circuit architecture unchanged) across all genomes producing maximum fitness circuits in each generation. In Figure 8, we show the increase in this sub-population's robustness over evolution for one GA run.

To determine if there exists a selective pressure for this pattern of increasing robustness to mutations, we performed the neutral mutation count on all of the forty-one GA runs that produced genomes resulting in maximum fitness. There is a highly significant difference in the number of neutral mutations in the best genome at the generation when the GA run first reaches a maximum fitness (and continues unbroken until the last generation) and at the last generation ( $P=0.000685$ , one-tailed Fisher Sign test).

To explore phenotypic variability, we determined the number of unique networks produced through all point mutations for the best individual at the beginning of the maximum fitness continuation and for the best individual at the last generation for all forty-one runs. The results of

this analysis are show in Table 1. These data indicate that buffering of development against genetic perturbations also decreased the phenotypic variation.

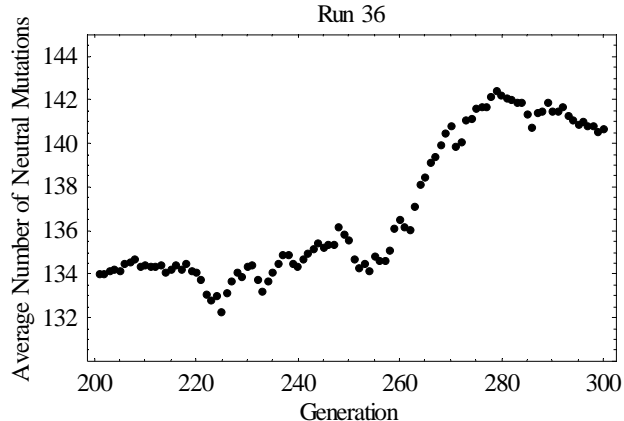


Fig. 8. Plot of the average number of neutral mutations for all maximum fitness individuals over the generations at which at least one genome achieved maximum fitness and remained unbroken over time.

Table 1. Phenotypic variability in canalized and non-canalized GA simulations

GA simulations	N	Variation		
		Decreased	Increased	Unchanged
Canalized	27	19	4	4
Non-canalized	14	1	11	2

N, number of runs

### Phenotypic Variation after Gene Knockouts

Bergman and Siegal (2003) show an increase in phenotypic variability after gene knockouts in their regulatory gene networks. For the canalized runs in our simulation, we deleted each gene in turn from the best genome at the last generation and then determined the phenotypic variation. All but one of the twenty-seven canalized runs remained unchanged or decreased in variation for the gene knockouts. The unique run increased its phenotypic variation by one network after knocking out a gene.

## Conclusion

Our work demonstrated that a process of target cell identification encoded into a genetic regulatory network can generate the synaptic connectivity of a known biological neural network and produce wide variation in neural network architectures through single-bit mutations. Regulatory networks provide combinatorial expression of developmental controlling factors leading to large phenotypic variation, the substrate upon which selection acts (Carroll et al. 2001).

But developmental pathways can become buffered against genetic and environmental disturbances thereby reducing phenotypic variation. In our simulations, after the maximum fitness had been obtained, genomic changes occurred that increased a GRN's ability to withstand mutations yet decreased the number of different networks produced through low genetic variation.

Phenotypic variability could not be regained through larger genetic perturbations in contrast to Bergman and Siegal's work (2003). This suggests a lack of regulatory complexity. Increased complexity can be achieved through two common regulatory motifs: gene repression (Hanna-Rose and Hansen 1996) or multi-component promoters (Lee et al. 2002), which were present in Bergman and Siegal's model. We plan to add gene repression next to the model and continue our investigation of phenotypic variability. Gene repression may allow for a release from canalization and greater phenotypic variation after gene knockouts.

Eventually, instead of using a topological fitness measure, we will implement behavioral fitness measures to explore the phenotypic variability necessary for the origin of behavioral novelty through connectivity changes.

## Acknowledgements

We thank Jennifer Talley and four anonymous reviewers for useful comments. This research was made possible in part by a NSF IGERT fellowship in Neuromechanics at Case Western Reserve University and by NSF grant EIA-0130773.

## References

- Bergman, A. and Siegal, M.L. (2003) Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424:549-552.
- Carroll, S.B., Grenier, J.K., and Weatherbee, S.D. (2001) *From DNA to Diversity*. Walsworth Publishing Company.
- Dellaert, F. and Beer, R.D. (1996) A developmental model for the evolution of complete autonomous agents. In P. Maes, M. Mataric, J. Meyer, J. Pollack and S. Wilson (eds.), *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. MIT Press.
- Edwards, J.S., and Palka, J. (1991) Insect neural evolution – a fugue or an opera? *Semin. Neurosci.*, 3:391-398.
- Fleischer, K. (1995) A multiple-mechanism developmental model for defining self-organizing geometric structures. Ph.D. diss., Dept. of Computer Science, California Institute of Technology.
- Hanna-Rose, W., and Hansen, U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* 12:229-234.
- Lee, T.I., et al. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799-804.
- Magoski, N.S. and Bulloch, A.G.M. (1998) Trophic and contact conditions modulate synapse formation between identified neurons. *J. Neurophysiol.* 79:3279-3283.
- Nishikawa, K.C. (1997) Emergence of novel functions during brain evolution. *Bioscience* 47:341-354.
- Nishikawa, K.C., Anderson, C.W., Deban, S.M., and O'Reilly, J.C. (1992) The evolution of neural circuits controlling feeding behavior in frogs. *Brain Behav. Evol.* 40:125-140.
- Rutherford, S.L. and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution. *Nature* 396:336-342.
- Shaw, S.R. and Meinertzhagen, I.A. (1986) Evolutionary progression at synaptic connections made by identified homologous neurons. *Proc. Natl. Acad. Sci. USA* 83:7961-7965.
- Spencer, G.E., Lukowiak, K., and Syed, N.I. (2000) Transmitter-receptor interactions between growth cones of identified *Lymnaea* neurons determine target cell selection *in vitro*. *J. Neurosci.* 20:8077-8086.
- Sperry, R.W. (1963) Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl. Acad. Sci. USA*. 50:703-710.
- Syed, N.I., Bulloch, A.G.M., and Lukowiak, K. (1990) In vitro reconstruction of the respiratory central pattern generator of the mollusk *Lymnaea*. *Science*. 250:282-285.
- Waddington, C.H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150:563-565.