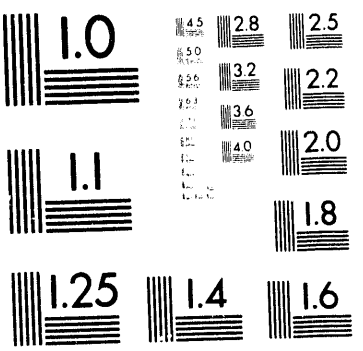
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

1 of 1

Conf-9306231--1

LA-UR 93-2827

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE: MODELING THE CONNECTION BETWEEN DEVELOPMENT AND EVOLUTION: Preliminary Report

AUTHOR(S): E. Mjolsness, Yale University
C. Garrett, University of Washington
J. Reinitz, Yale University
D. Sharp, T-13

SUBMITTED TO: The Proceedings of the Monterey Bio-Computations Conference held June 21 - 28, 1993 in Monterey, CA.

DISCLAIMER

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

FORM NO 836 R4
ST NO 2629 5 81

Modeling the Connection between Development and Evolution: Preliminary Report

Eric Mjolsness*
Charles D. Garrett†
John Reinitz*
David H. Sharp‡

*Department of Computer Science, Yale University
P.O. Box 2158 Yale Station, New Haven CT 06520-2158
†Department of Computer Science, University of Washington
Seattle, WA 98105
*Center for Medical Informatics, Yale University
New Haven CT 06510
and Brookdale Center for Molecular Biology, Mt. Sinai School of Medicine (present address)
1 Gustave L. Levy Place, New York NY 10029
‡Theoretical Division, Los Alamos National Laboratory
Los Alamos NM 87545

July 29, 1993

Abstract

In this paper we outline a model which incorporates developmental processes into an evolutionary framework. The model consists of three sectors describing development, genetics, and the selective environment. The formulation of models governing each sector uses dynamical grammars to describe processes in which state variables evolve in a quantitative fashion, and the number and type of participating biological entities can change. This program has previously been elaborated for development. Its extension to the other sectors of the model is discussed here and forms the basis for further approximations. A specific implementation of these ideas is described for an idealized model of the evolution of a multicellular organism. While this model does not describe an actual biological system, it illustrates the interplay of development and evolution. Preliminary results of numerical simulations of this idealized model are presented.

1 Introduction

The scientific questions motivating the present work concern the many instances in which development and evolution interact. For example, it has been hypothesized [1] that the period of experimentation with alternative body plans (the "Cambrian explosion") just after the evolution of multicellular organisms ended because of an interaction between developmental

and selective constraints. Since basic animal body plans have remained fixed since that time, a theoretical understanding of this problem is of great interest. A model of the evolutionary dynamics of body plans requires development as an essential component. More generally, the expected diversity of morphologies in a given environment is an open scientific question which leads to the study of the interaction between development and evolution.

Another fundamental question about evolution concerns the biological units on which selection acts. In the Modern Synthesis, multicellular organisms are taken to be the units of selection, but in the early history of life cells and even smaller units played this role. It is possible that the ancient transition from cellular to multicellular units of selection must be modeled in some detail in order to understand aspects of multicellular organization such as the sequestration of the germ line and common patterns of gastrulation in animals. Since the cellular and multicellular levels are related through development, a model of the transition between units of selection will necessarily require a model which integrates developmental and evolutionary processes.

It is the purpose of this paper to formulate a model describing the interaction between development and evolution, and to present the results of preliminary computer simulations. These simulations illustrate schematically the transition from reproducing cells to reproducing multicellular organisms. The model proposed here is an extension of our previous work on development [10, 12].

Developmental and evolutionary processes occur at different levels of organization such as genomes, cells, cell clusters, organisms, populations of organisms and ecosystems. Important interactions occur within and between these levels of organization. This presents a multiscale modeling problem of daunting complexity, but we believe that a tractable approach can be based on the fact that the problem decomposes naturally into modules, or sectors, which can be developed and modified more or less independently. In particular, the relevant processes can be organized conceptually into model sectors describing development, the selective environment, and genetic adaptation through heritable variation.

Existing models of evolution treat the map from genotype to phenotype in a rudimentary way. A more complete modeling framework for development was introduced in [10]. This model was based on a set of ordinary differential equations (ODEs) to represent a genetic regulatory circuit, together with a set of rules (a "grammar") which allow a description of changes in the number and type of biological entities that are present at a given time. We will adopt this model to describe the developmental sector of the enlarged model of evolution discussed here, and we will also borrow its techniques in formulating models for the other sectors.

A key issue in formulating the sector of the model pertaining to the selective environment is to define a scoring function which reasonably approximates a "fitness function". We are aware that the precise character of such a fitness function, in fact even its existence, remains an open scientific question in spite of much effort. We view a fitness function as an approximation to a more complete model of a selective environment, which would use a stochastic grammar to describe important interactions with the environment such as establishing a territory, finding a mate and getting killed. The formulation of such a model would use dynamical grammars in a way similar to their use in the developmental sector, but

is not attempted here. The work in the present paper is based on a simple scoring function which accounts for several aspects of the selective environment for multicellular organisms. (See Section 2.3).

We next discuss the genetic sector. Important genetic events to be modeled include point mutations, cross-over, gene duplication and fusion of gametes. An accurate description of the impact of these events on a genetic regulatory circuit is another open scientific problem. As a substitute for this genetic sector, we use a crude model of adaptation based on simulated annealing. It differs from real genetics in that the only genetic events are "point mutations" acting directly on the regulatory circuitry, which are selected using an optimization algorithm [5, 6, 7]. In this paper we restrict attention to asexually reproducing haploid organisms.

This paper also reports results of preliminary numerical experiments with the model. In these experiments, we imagine a hypothetical world as follows. As an initial state, we have a collection of cells which can reproduce to form multicellular aggregates. These aggregates, however, do not have a genotype capable of multicellular reproduction. The cells' genotype is allowed to adapt under a fitness function which rewards multicellular organisms of a certain size and internal structure, and which reproduce as multicellular entities. In this world, development is specified by a simple genetic circuit model. We exhibit simulations of such a system, and examine the various outcomes as to their capacity for multicellular reproduction and the behavior of their associated genetic circuits.

Experience with the developmental sector of the model as applied to the *Drosophila* blastoderm has taught that the modeling approach described here can be successfully applied to answer real biological questions, but only when great care is exercised in incorporating the necessary biological detail. The same will certainly be true of the evolutionary extension of the model outlined here. The computer experiments described serve to illustrate the interplay of development and evolution in the model, but do not yet constitute a biologically realistic model of an evolutionary process.

2 Formulation of the Model

2.1 Development

We have described our modeling framework for development elsewhere [10]. We recapitulate it here only to the extent necessary to understand its application in the present paper. As used here, the model describes two fundamental processes: (a) the continuous dynamics of regulatory molecules and (b) the change in number and type of biological entities.

Concentrations of regulatory molecules change in response to existing concentrations of regulators, exchange of regulatory molecules between nuclei (by diffusion), and decay. These effects are described in the model by a system of coupled nonlinear differential equations. The continuous internal dynamics of gene products are given by a "genetic regulatory circuit" in which a pair of genes a and b interact by means of a single real number T^{ab}. There may be many proteins b regulating the gene for protein a and thus influencing the dynamics of its concentration v_i^a in object i. To make a tractable model, we will assume that these effects are monotonic in the concentrations v_i^b and are approximately additive, with nonlinearity

confined to sigmoidal threshold functions g_a. So we assume that the dynamics of v_i^a depend on the other variables v_i^b through a summed input u_i^a:

$$u_i^a = \sum_b T^{ab} v_i^b + h^a, \tag{1}$$

where h^a determines the threshold of g_a. During the time when object i does not participate in birth or death processes, which for dividing cells is called interphase, we use a continuous time model of its internal dynamics. A simple model is the neural net dynamics given by (c.f. [4])

$$\tau_a \frac{dv_i^a}{dt} = g_a(u_i^a) - \lambda_a v_i^a. \tag{2}$$

Equations (1) and (2) comprise our connectionist model of the internal dynamics of interphase, one of many processes involved in development. Other continuous time and purely internal dynamical subsystems will be modeled in like manner, with just the connection matrix $\mathbf{T}$ and the thresholds h^a changed. Each such process can be represented by one rule in a dynamical "grammar" which can model several biological processes.

We next consider discrete events. Birth processes such as cell division (mitosis) require a discrete time update equation with its own connection matrix. Since the different daughter cells of one parent cell are not necessarily equivalent, we use one such connection matrix $\mathbf{T}_k$ (with components T_k^{ab}) for each of the progeny. We use multiple index notation: if i is the index of the parent cell, then (i,k) is the index of its k'th daughter cell. (And $((i,k),l)$ would index the second generation descended from i.) We then suppose that

$$v_{(i,k)}^a = v_i^a + R_a g_a(\sum_b T_k^{ab} v_i^b + h^a) - \hat{\lambda}_a v_i^a. \tag{3}$$

When the only regulatory molecules are gene products, Equation (3) can be further simplified. Because there is no synthesis of gene products during mitosis, the only dynamical process that modifies v is unequal partioning of gene products, and Equation (3) becomes

$$v_{(i,k)}^a = U_k^a v_i^a \tag{4}$$

where each U_k^a is diagonal and $U_k^a \geq 0$. Other discrete-time processes, which just change the type and state of one object, can be modeled as one-child analogs of Equations (3) or (4) by suppressing the k index.

For a system which includes both interphase and cell division, Equations (2) and (3) must be combined so that (2) operates continuously except at certain discrete times when (3) is invoked. The same is true of any combination of continuous and discrete time processes.

We see that every developmental process that we model is described by a set of rules which govern how a single object of a given type can be replaced by one or more objects of the same or different type, together with an internal dynamics model such as that defined by Equations (2) and (3). This set of rules can be thought of as a developmental "grammar" Γ, in the sense of Lindenmayer and others [8, 2, 11]. The grammatical rule adopted by an object i at a given time t will in general be a function of its state vector $\mathbf{v}_i(t)$. In summary, this approach comprises a framework in which many biological processes, continuous and

discrete in time, can be modeled in a unified way. One can take advantage of this fact to model different processes at different levels of detail. This may be of considerable utility for multiscale modeling in biology.

As an example, we discuss a five-rule grammar for a collection of cells that have the capability of dividing and of forming reproducing multicellular organisms. These rules will be used in the experiments reported in Section 3. We assume that the internal state of each cell is specified by the concentrations of a small set of proteins that are products of regulatory genes, and in particular that these gene products control cell division. The cells do not exchange material with one another, but they do pass material to their progeny. Although mitosis itself is a complex dynamic process, we model only those aspects relevant to the problem at hand: Mitosis lasts a finite time, during which genes are not expressed, and at the end of that time there are two cells instead of one. In order to describe a multicellular organism, there must be a rule that represents the idea of the temporal boundaries of such an organism. Lastly, we must take into account the fact that not all zygotes survive to find an available territory (or more generally a "niche" or "slot") in the environment (see Section 2.3) This picture can be summarized in the following five rules.

1 cell $\rightarrow$ cell, or interphase, in which the internal state variables of the cell evolve continuously;

2 cell $\rightarrow$ mitosing cell, which initiates cell division. During mitosis the genetic circuitry is shut down for a period which represents the time required for a cell's chromosomes to condense and separate on the mitotic spindle;

3 mitosing cell $\rightarrow$ two cells, or cleavage, which concludes mitosis;

4 cell $\rightarrow$ (detached) cell, or dispersal, in which a zygote separates from its parent organism.

5 (detached) cell $\rightarrow$ cell, in which the zygote successfully starts a new organism.

2.2 Genetics

Phylogeny is a summary of the evolutionary history of an organism. Darwinian evolution consists of selection, discussed in the next section, and heritable variation, discussed here.

A phylogeny can be pictured (Figure 1) as a comprehensive lineage tree, showing the ancestry of every cell, which has three or four scales of organization: (1) individual cells and their descendants, (2) some number of unit subtrees, each of which recurs approximately across generations and constitutes the ontogeny of a phase in the life cycle of an organism, (3) instances of the full life cycle of an organism, and (4) phylogeny: the slow emergence over evolutionary time of a repeating life cycle and its constituent subtrees. Actually, the tree should be generalized to a directed acyclic graph (DAG) because some cells receive information from more than one "parent" during recombination, the fusion of gametes, or developmental induction. For the present we will ignore these important processes, and the genetic effects of diploidy, and we will also indirectly but severely limit the possibilities for multiple-phase life cycles (thereby conflating scales (2) and (3) above) by modeling only

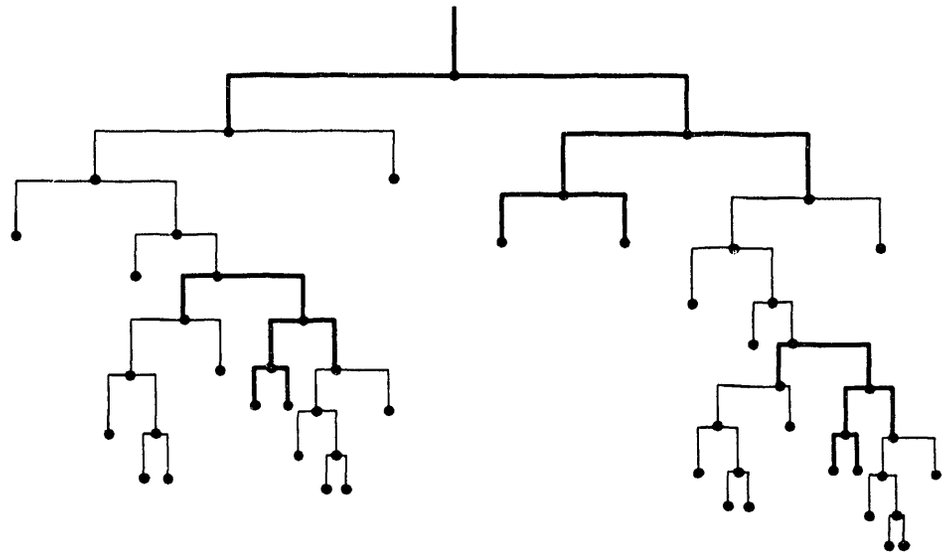

Figure 1: A comprehensive lineage tree, showing organization at different scales. Cells (nodes) are grouped into two recurring subtrees (heavy and light lines). One heavy and zero, one, or two light subtrees make up one organism. Organisms repeat in a large-scale lineage.

very limited environments. Thus we will model comprehensive lineage trees with cellular, individual, and phylogenetic levels of organization.

If we examine the genotype of each cell in such a comprehensive lineage tree, we will see that cells and their progeny often differ owing to genetic events which introduce heritable variation. These include point mutations, gene duplications, and crossover. (Technically, crossover and sexual reproduction change the tree into a DAG.) Such processes may be modeled as additional grammatical rules which modify a cell's genotype. By contrast, the developmental grammar rules we have considered previously act on the protein concentration state vector. Thus one has a combined grammar, which has both developmental and genetic rules.

We note that selective processes are not represented directly in the comprehensive lineage tree, but their result (mortality of an organism) is reflected in changes in the shape of the tree. This occurs since fewer progeny are included in the tree.

Full implementation of these ideas requires information that we do not presently have. Specifically, the action of genetic grammar rules (such as those describing point mutation and crossover) on genetic circuits requires a biologically realistic model of the relationship between genetic information on a one-dimensional string of DNA and the connection matrix of a genetic regulatory circuit.

As an interim measure we take advantage of a similarity between simulated annealing and the comprehensive lineage tree with genetic events to formulate a rough approximation

to the genetic dynamics. Simulated annealing is a method for finding the global minimum of a "bumpy" function, that is, one with many local minima. The method is derived from statistical mechanics [9] where it models the slow cooling (annealing) of a physical system to its lowest energy state; later it was generalized by Kirkpatrick [5]. In the following, the function to be minimized (the "cost function") is $E = f(x_1, \ldots, x_i, \ldots x_n)$, and T is a parameter (the "temperature") that starts off large and slowly gets smaller. The basic method is quite simple. Start with a random set of x_i, and then perform the following procedure:

1. Compute $E = E_{\text{old}}$ from the variables x_i.
2. Make a change in one (or more) of the x_i (This is referred to as a "move").
3. Compute $E = E_{\text{new}}$ from the newly generated set of x_i.
4. Compute $\exp\left(\frac{E_{\text{old}} - E_{\text{new}}}{T}\right)$
5. If the above quantity is bigger than a random number between zero and one, keep the new x_i's ("accept the move"). Otherwise, restore the old x_i's ("reject the move").
6. Repeat while allowing T to decrease slowly from a large value to zero. Typically this entails 10^5 to 10^7 iterations.

Our simulated annealing algorithm [6, 7] consists of point mutations on connection matrix elements and rule strength coefficients, selection using a scoring function described in the next section, and gradual reduction of both the temperature and move size to make an optimization algorithm. Point mutation on connection matrix elements is directly analogous to point mutation on genetic information, and stochastic update using a scoring function is directly analogous to stochastic selection using a fitness function. However, the regulation of temperature and the reduction in move size over the course of an annealing run does not have a direct analogy in the comprehensive lineage tree picture. They represent, in essence, a stochastic process approximating such a picture. This approximation achieves computational efficiency in that the effects of macromutations are explored under low selective pressure. As the selective pressure increases with the decline in temperature, the effects of mutations with smaller and smaller effects are explored. These points of analogy are exploited in the algorithm used in the computer experiments reported in section 3.

2.3 Selective Environment

As in development, a dynamical model of a selective environment must include diverse processes, many of which change the number of organisms that are present. We have already mentioned mating and various forms of mortality as examples of such processes. For this reason, the use of a grammar in which a rule corresponds to a process again appears to be a useful modeling tool. The grammars used in modeling the selective environment will differ from the developmental grammars in that they are stochastic grammars, reflecting the fact that most environmental processes have an important stochastic component, and that such

processes are often too intricate to model deterministically. As in developmental grammars, however, we expect that there will be a quantitative dynamics associated with each process.

The grammar rules to be implemented in a specific problem must reflect a careful analysis of the features in the environment which are of central importance. Given a scientifically correct identification of the rules to be modeled, there is still the question of whether the full probabilistic description can be simplified.

The exact probabilistic formulation amounts to a master equation for the probability distribution of numbers of organisms of each species and phenotype. We consider coarse-grained phenotypes, so that the average occupancy for each phenotype can be high in a large population. Then we can apply the van Kampen–Kramers–Moyal [3] "large system" approximation to the master equation to obtain simplified stochastic dynamics in the form of Fokker-Planck equations. These equations have as their solutions Gaussian probability distributions, with time dependent mean and covariance which approach well defined asymptotic limits. We suggest that the limiting Gaussian may provide the defining distribution in which to evaluate the expected number of offspring, or closely related quantities such as reproductive value, and hence in which to calculate the fitness function for a given species. Indeed, fitness so calculated may have the same maximum as the limiting Gaussian. It is in this sense that we think of a fitness function as approximating the stochastic grammar which models a selective environment.

This point of view raises several interesting questions. First, how well is the science of this problem captured by the Fokker-Planck approximation? If this approximation is not good enough, what are the prospects either for better analytic approximations (eg. higher moments) or for direct numerical solution of the master equation? Second, in a compound grammar modeling several interacting species, how are the separate fitness functions related to the asymptotic Gaussian solution of the entire system? An answer to this question could shed interesting light on the game theory approach to evolutionarily stable strategies [13], because it leads naturally to an analysis of the asymptotic stable states of a system with competing species. As a third question, we ask how to treat fine-grained phenotypes with average occupancy numbers $\ll 1$.

We believe that fitness functions may provide useful approximations to the dynamics of the selective environment in some circumstances. We shall assume that this is the case for the model studied in Section 3 of this paper, and go on to describe a further approximation of the fitness function by a scoring function.

The main idea of this further approximation is to represent several selective effects as additive terms in a scoring function. This scoring function is *ad hoc* in that its separate terms are guesses as to approximate analytical forms for various contributions to the true fitness function; they have in no sense been derived or computed.

The environmental scoring function rewards multicellular organisms which develop an optimal number of cells, correctly distributed into three cell types, achieved through a supportable rate of proliferation. This much results in ontogenies. Life cycles result if the scoring function is actually a sum of such functions, taken over a set of available environmental *slots* each of which may be occupied by at most one individual. The occupation of such slots is governed by further environmental parameters, and requires that some cells of established

organisms *disperse* to another slot. These processes can be represented analytically with a scoring function of the form

$$S \simeq \sum_{\text{slots}} \sum_{\text{time}} \left\{ (1/2)(\frac{n}{n_{\text{target}}} - 1)^2 + (1/2) \sum_{\text{cell types } \alpha} (\frac{n_\alpha}{n} - f_{\alpha \text{ target}})^2 + \frac{\Delta n_{\text{mitosis}}}{n^\gamma} \right\} \tag{5}$$

($\gamma = 2/3$ or 1). In this formula, n is the number of cells present in a given environmental slot at a given time, n_α is the number of cells which have differentiated to cell type α, and $\Delta n_{\text{mitosis}}$ is the number of cells produced by mitosis at a given time step. The constants appearing in this formula are n_{target}, the desired number of cells in an individual organism, $f_{\alpha \text{ target}}$, the desired fraction of cells differentiated to type α, and the exponent γ which relates the energetic cost (or cost in some other limited resource) of each mitosis to the energy or resource available within the organism. The constants n_{target} and $\{f_{\alpha \text{ target}} | \sum_\alpha f_{\alpha \text{ target}} = 1\}$ can be specified arbitrarily, subject to practical constraints on the cost of the simulations.

The first two terms in the scoring function are simply quadratic penalty functions which favor organisms which have parameter values close to those desired.

The last term represents a sharing of energy or some other limited resource among all the cells in an organism. to support mitosis by a few. The form of this term is the fraction of the organism's limited resource expended in mitosis. The denominator is the total resource available, and is a function of the size of the organism, n. This function could express a proportionality to the volume ($\gamma = 1$) or to the surface area ($\gamma = 2/3$), for example. The mitosis term thus incorporates the essential assumption that the cells in a slot are actually cooperating and functioning as a multicellular organism. The first two terms also reflect cooperation, but in the weaker sense that separate cells produce the correct number and types of offspring whether or not they interact during development.

The summations in (5) express the fact that a genotype is shared by all the cells in an organism, over several generations. A genotype is therefore scored not only by its ability to fill one slot well, but also by its ability to occupy new slots. Finding and occupying a slot is governed by important environmental parameters such as the total number of slots present at a given time and the the probability of finding and occupying an unoccupied slot (as a function of time). In our simulations we choose a schedule of slot creation, destruction, and probability of occupancy which is periodic, as shown in Figure 2. This choice of schedule can be interpreted as favoring the evolution of "iteroparous" as compared to "semelparous" organisms. These two classes of organisms differ in their reproductive strategies. Iteroparous organisms produce offspring in a slow, steady manner (eg. horses), whereas semelparous organisms produce many offspring in a rapid burst of reproductive activity (fruit flies). These reproductive strategies are associated with low variance and high variance environments respectively.

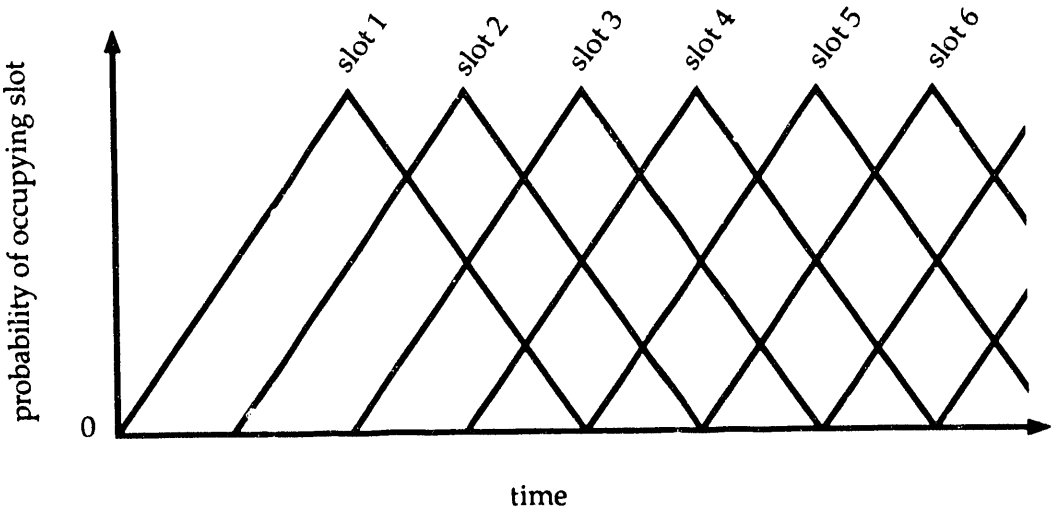

Figure 2: Assumed probability with which a zygote may find and occupy an unoccupied slot as a function of time. Different slots are temporally overlapped as shown. Each slot's probability of being suitable for a new zygote (if the slot is empty) starts at zero, rises to a maximal value $p_{max} \leq 1$, then returns to zero. Experiments used $p_{max} = .5$. When the probability returns to zero the slot is eliminated and its occupant dies.

3 Results of Numerical Computation

In this section we complete the specification of the model which we have simulated (Section 3.1), describe how the simulations were performed (Section 3.2), and (Section 3.3) report and discuss the results.

3.1 Specification of the Model

Here we recapitulate briefly the components of our model, as described in the previous section.

The growth and reproduction of a multicellular organism is modeled by a dynamical grammar (Section 2.1). This grammar makes possible the growth and reproduction of multicellular organisms by incorporating rules governing cell division, gene regulation, and dispersal of zygotes. However, the grammar does not specify how these events are coordinated.

The selective environment is modeled by a scoring function (Equation 5) which rewards reproductive multicellular organisms of a certain size and degree of cellular differentiation.

Finally, we have a genetic sector whose function is to make changes in the genome which by trial and error optimize the environmental scoring function. The parameters which are varied to achieve the optimization include connection strengths in the genetic regulatory circuits, parameters governing cell type differentiation, and the parameters governing the application of discrete time grammar rules. The trial and error procedure is implemented using the method of simulated annealing.

3.2 How the Computations Were Performed

We mentioned that in broad outline the computations were performed by applying the method of simulated annealing to optimize Equation (5). In this section we explain further details of this procedure.

The simulated annealing algorithm used was a variation of the one proposed by Lam [6, 7], in which both the temperature and average step size are controlled according to an adaptive cooling schedule. Annealing moves for each parameter were selected from an exponential distribution and given random signs. The mean of this distribution was separately altered for each parameter so as to ensure an effective sampling of the search space. This search space was restricted by limiting the saturation of the threshold function $g(u)$, Equation (2), to 99%.

The parameters to be optimized characterize genetic circuits and their connection to grammar rules. There is a genetic circuit associated with each grammatical rule in section 2.1. Rule 1 (the interphase rule) has a fully connected matrix defined by Equation 1. The remaining rules 2 through 5 operate in discrete time, and of these rules, only the matrix elements in rule 3 (cleavage) and rule 5 (formation of a new organism) are optimized. Rules 2, 3 and 5 have discrete-time diagonal circuits, specializing Equation 3 (or its one-child analog in which the k index is suppressed) to the case in which T is diagonal. Rule 2 (mitosis) has a constant $T = 0$. Rule 4 (dispersal) does not affect the internal state vector at all ($U^a = 1$ in the one-child analog of Equation 4).

Further parameters are required to control the processes of rule selection and cell type differentiation which we imagine summarize the effects of other regulatory circuits not explicitly included in the model. If such an omitted subcircuit has one input unit, and perhaps many output units, then the net input to that circuit from the explicitly modeled regulatory circuit would be, as in Equation 1, a function of the vector product between a row of a connection matrix and the state vector of the regulatory circuit plus a threshold. So for each grammatical rule $r \in \{1 \ldots 5\}$, we assume a rule strength of the form

$$S_i^r = \mathbf{s}^r \cdot \mathbf{v}_i + \theta^r \tag{6}$$

We also assume that the rules compete at each cell, and that the one with the largest strength is selected.

Many of the strength connections $\mathbf{s}^r$ and θ^r are made identically zero in order to impose appropriate structure on the system. Rule 1 (interphase) is the only continuous time rule, with $\mathbf{s}^{int} = 0$, and has a strength $\theta^{int} = 1$ which is independent of $\mathbf{v}_i$. Rules 2 (mitosis) and 4 (dispersal) have $\theta^r = 0$. Typically one of these rules is triggered when its strength rises above θ^{int}. $\mathbf{s}^{mitosis}$ was held constant, while $\mathbf{s}^{dispersal}$ was optimized. Rule 2 (mitosis) always triggered rule 3 (cleavage) on the next time step, and rule 4 (dispersal) triggered rule 5 (new organism) in a stochastic manner according to the probability distribution shown in Figure 2. Failure to trigger rule 5 resulted in death.

Similarly, cell type differentiation is determined at each cell type i by a competition between cell type strengths given by the linear form

$$c_{\alpha,i} = \mathbf{c}_\alpha \cdot \mathbf{v}_i. \tag{7}$$

In this case a "soft" competition was implemented in which for small cell type strengths $c_{\alpha,i}$ a cell can belong partially to different cell types. Then the number of cells of type α appearing in the scoring function is given by

$$n_\alpha = \sum_{\text{cells } i} \left[e^{\mu c_{\alpha,i}} / \sum_\beta e^{\mu c_{\beta,i}} \right]^2. \tag{8}$$

The square encourages decisiveness but does not absolutely require it. The cell type strength vectors, $\mathbf{c}_\alpha$, were also optimized.

Experiments were performed with and without "slots", i.e. with selection for multicellular growth but with and without selection for multicellular reproduction. Experiments with slots used a sum over twelve slots in the scoring function (5). The procedure for experiments without slots differed from the foregoing in that the dispersal grammar rule was omitted and the scoring function was not summed over slots. Furthermore, the 5×5 interphase circuit was assumed to have the special form of non-communicating 2×2 and 3×3 sub-circuits. This type of experiment was done as a pilot experiment to explore multicellular growth.

A typical simulated annealing run required 2 to 4 days of CPU time on a SPARC 2 workstation.

3.3 Results

In this section we report the results of numerical simulations. We emphasize that these results are preliminary, in that we present the outcomes of individual simulated annealing runs. We have yet to characterize exhaustively the possible behavior and lineages which can arise from the chosen scoring function.

In the first set of runs, the optimization was performed without summing the scoring function over slots. Several runs produced multicellular organisms having the desired size, distribution of cell types, and growth schedule. A cell lineage tree for one such run is shown in Figure 3. The desired number of cells, 24, was not a power of 2 and thus an unbalanced tree was required to achieve the correct size. The strategy adopted to arrive at this unbalanced tree depended on differential reproductive rates in the left and right progeny cells, which came about because cells on the right branches of the tree stopped reproducing sooner. We refer to this loss of mitotic capacity as "aging". The mechanism that causes aging can be seen in the phase portrait shown in Figures 4 and 5. Figure 4 shows the convergence of the interphase dynamics alone to a distant fixed point, regardless of the initial conditions. Figure 5 shows a small region of Figure 4, superimposed on the full dynamics of the growing organism. Cells begin mitosis when their interphase dynamics takes protein 0 above its threshold concentration. For sufficiently large concentrations of protein 1, this can no longer happen: protein 0 can not reach threshold because the interphase trajectories bend towards the distant interphase fixed point. Therefore, increased levels of protein 1 result in a termination of a cell's ability to reproduce. Left and right progeny age differently, as seen in the figure, because the cleavage grammar rule affects the concentration of protein 1 differently for the two progeny (see Equation (4)).

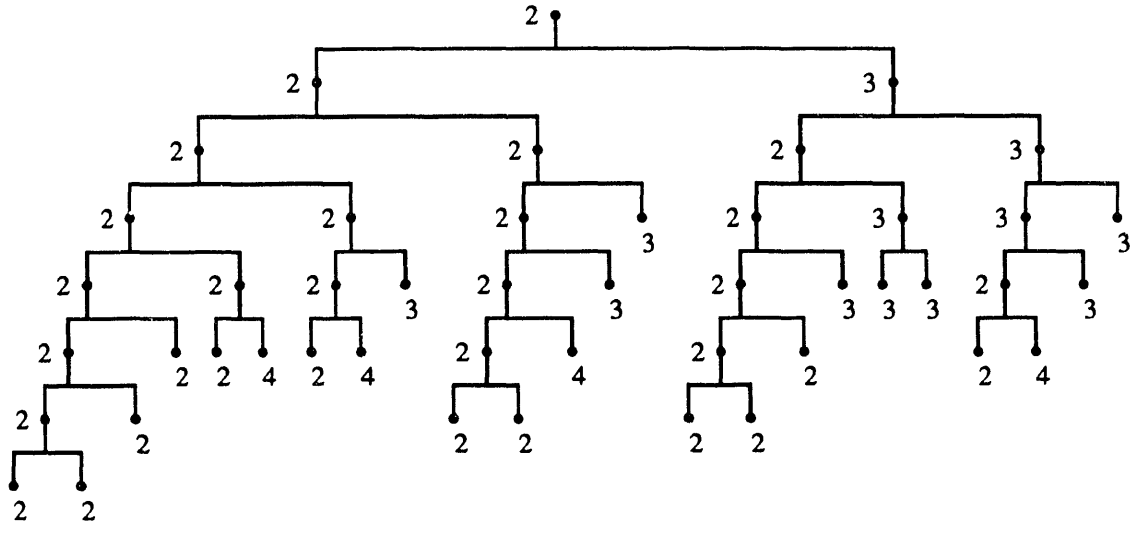

Figure 3: An evolved lineage tree for a single-slot organism (no reproduction required). Cells are labelled by their scoring function cell type. The desired numbers of cells of each type are attained: there are 12 terminal cells of type 2, eight of type 3 and four of type 4.

In the same run, cellular differentiation occurred. The desired fractions $f_{\alpha\,\mathrm{target}}$ of the three different cell types were 1/2, 1/3 and 1/6. Figure 6 is a phase portrait of two components of the 3×3 sub-circuit, showing how eight of the cells are differentiated from the rest. These cells, which have low concentrations of protein 3, appear to be in the domain of attraction of a nearby fixed point, whereas 15 of the 16 remaining cells appear to be converging to a distant fixed point. In no case does a cell reach its interphase fixed point; the cells have been sufficiently separated to differentiate.

In the second set of experiments, the sum over twelve slots is included in the scoring function, so that a successful genotype must not only grow, but also reproduce as a multicellular organism. Figure 7 shows the lineage tree of our first exactly reproducing genotype. Cells are labeled according to whether they disperse or not, and by their cell type. Since dispersing cells are no longer part of the organism, the organism has the correct size.

The pattern of dispersing cells in the lineage tree can be understood from the phase portrait in Figure 8. This figure shows that the cells segregate into a sequence of 6 groups. Mitosis replaces a cell in group n with cells in groups $n+1$ and $n-1$. Dispersing cells are all found in group 6, except for two cells in group 4. These facts account for the pattern of dispersal in the lineage tree. The group number of each cell, along with its scoring function cell type and its status as a zygote or a somatic cell, is shown in Figure 8. We see that group 6 is specialized to produce zygotes, but other groups do not correspond in a one-to-one fashion with the externally imposed cell types. We may interpret the groups revealed in this figure as "emergent cell types", which are not required by the scoring function but are part of the developmental strategy which evolved to minimize the scoring function.

The growth and reproduction over many slots of organisms bearing the evolved genotype is shown in Figure 9. The number of cells in each organism is plotted against time, and

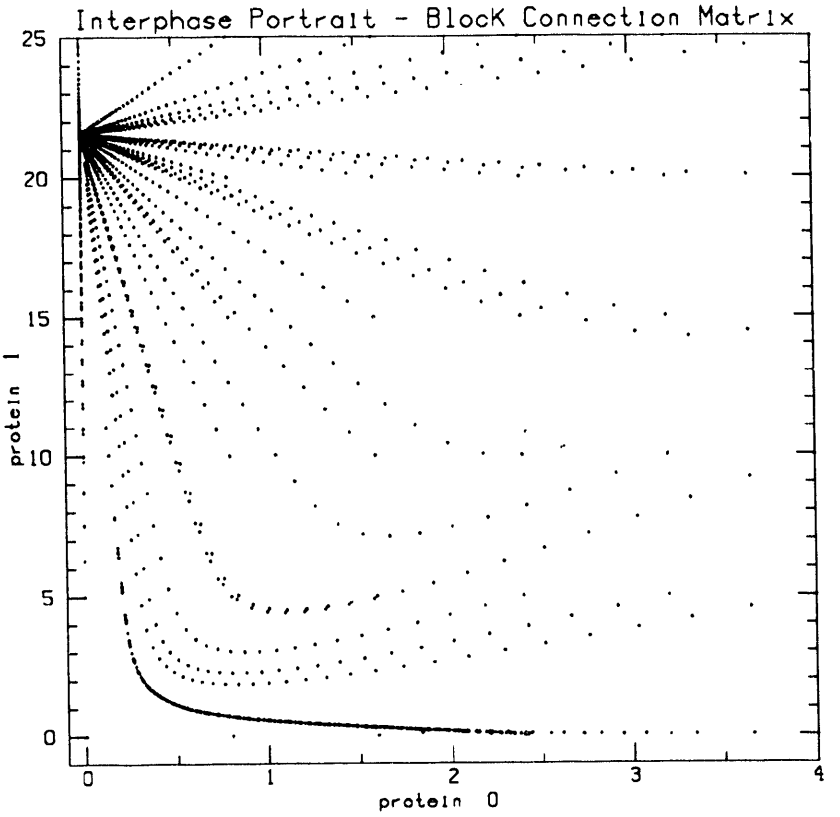

Figure 4: Evolved interphase dynamics for the single-slot organism. Observe fixed point at $(0, 21.5)$, and attracting trajectory from $(2.4, 0)$ to the fixed point.

the plots are superimposed. Starting with the appearance of the seventh slot, the growth patterns of the organisms are indistinguishable. We interpret this as an exact life cycle occurring in a stable environment.

The growth pattern in Figure 9 settles into an exact periodic life cycle after an initial transient behavior. This transient is an artifact of the use of simulated annealing as a substitute for the full genetic grammar discussed in Section 2.2. Under a genetic grammar, the initial state of a zygote's protein concentration vector after a change in genotype would be the same as the final state vector before such a change. But the simulated annealing procedure must re-set the state vector to some constant value for the zygote of the organism in the first slot, because the scoring function must be evaluated the same way after each genetic change.

After this transient, the run in Figure 9 has periodic behavior with period 1 in units of slot appearance time. Other periodicities are observed in other runs. Figure 10 shows the onset of a period 6 life cycle. On inspection this life cycle is seen to consist of 3 interleaved period 2 life cycles. As an indication of the range of results possible in this model, we mention that we have also observed in the simulations an apparently "chaotic" and non-terminating life cycle.

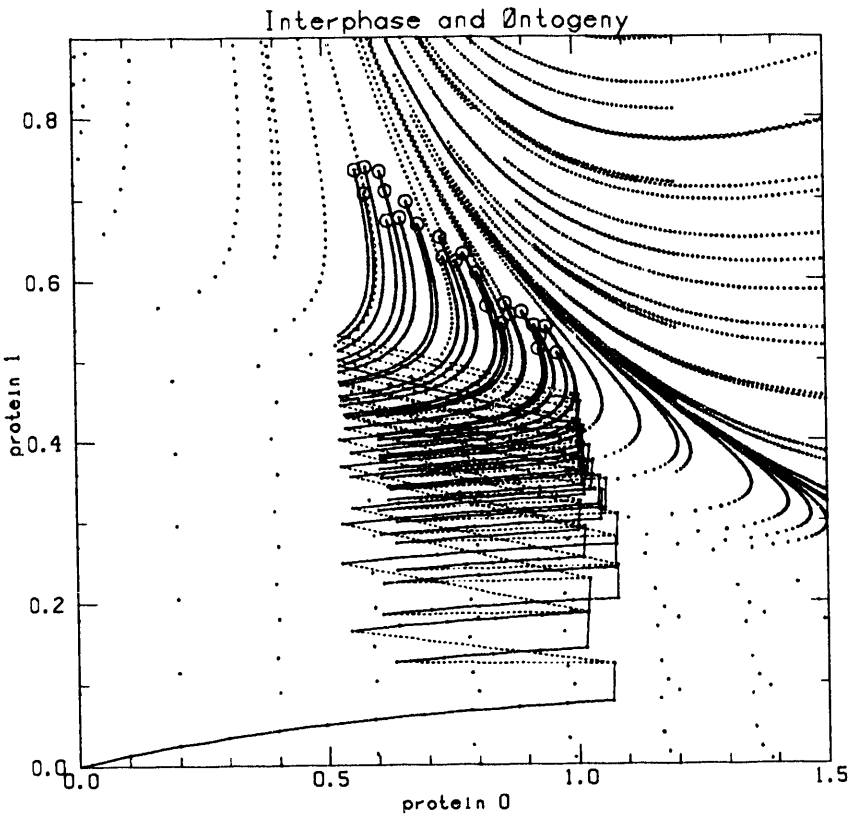

Figure 5: Phase portrait for the single-slot organism, showing an emergent "clock" which regulates the size of the organism. Note that the interphase fixed point is far outside the operating range of the genetic circuit.

4 Conclusion

We have presented a model incorporating developmental processes into an evolutionary framework. In simulations of this model, we have observed a number of interesting phenomena including an emergent clock which regulates organism size; the use of interphase attractors to define externally imposed cell types; emergent cell types which were not imposed by the environment; and multicellular reproduction in periodic and nonperiodic forms.

The next step is to identify an actual biological system for which the information required in the formulation of the model is currently available. The challenge is to apply the model to a real biological system, and thereby validate the correctness of the modeling assumptions and show that the model has scientific substance, in that it is capable of answering questions of biological interest. One area of study which may be suitable for these purposes has been suggested to us by Buss; this concerns the evolution of developmental differences between semelparous and iteroparous organisms, which must reproduce in relatively unpredictable and predictable environments respectively.

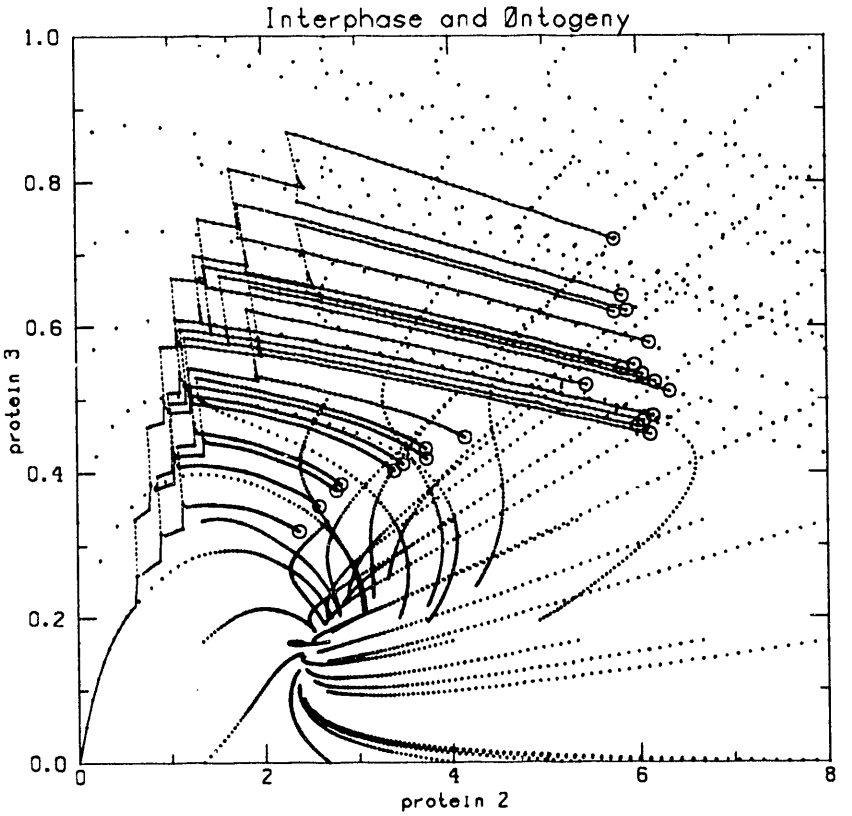

Figure 6: Another phase portrait for the single-slot organism, showing the differentiation of one third of the cells (those with the lowest values of protein 3). Interphase trajectories can cross because they have been projected from a three dimensional subcircuit to a two-dimensional plot.

Acknowledgement

We wish to thank Leo Buss for extensive discussions of the context and content of this research, and specifically for encouraging us to apply the developmental model to problems of biological evolution. JR acknowledges support from National Institutes of Health grants LM 07056 and RR 07801; EM and CG from the US Air Force Office of Scientific Research grant 88-0240; EM from Hewlett-Packard and from the Yale Institute for Biospheric Studies; and DHS from National Institutes of Health grant RR 07801.

References

[1] Leo W. Buss. *The Evolution of Individuality.* Princeton University Press, Princeton, New Jersey, 1987. See pp. 31, 69, 98-115.

[2] Kurt Fleischer and Alan H. Barr. A simulation testbed for the study of multicellular development: The multiple mechanisms of morphogenesis. In *Artificial Life III.* Addison-Wesley, 1993. To appear.

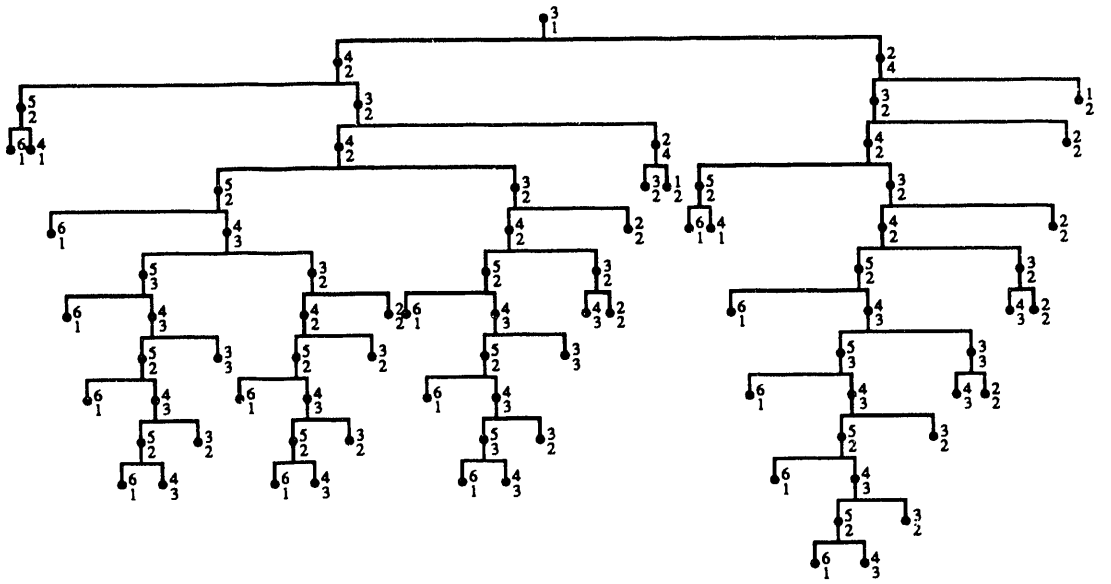

Figure 7: Lineage tree of the first exactly reproducing genotype. Cells are labeled above by their group number as determined from the next figure, and below by their status as zygotes or somatic cells (label = 1 for zygotes) and, for somatic cells, by their scoring-function cell type (label = 2, 3, or 4). Ontogeny of slot seven is shown.

[3] Crispin W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences.* Springer Verlag, Berlin, 1983.

[4] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences USA*, vol. 81:3088–3092, May 1984.

[5] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[6] Jimmy Lam and Jean-Marc Delosme. An efficient simulated annealing schedule: derivation. Technical Report 8816, Yale Electrical Engineering Department, New Haven, CT, September 1988.

[7] Jimmy Lam and Jean-Marc Delosme. An efficient simulated annealing schedule: implementation and evaluation. Technical Report 8817, Yale Electrical Engineering Department, New Haven, CT, September 1988.

[8] A. Lindenmayer. Mathematical models for cellular interaction in development, parts i and ii. *Journal of Theoretical Biology*, 18:280–315, 1968.

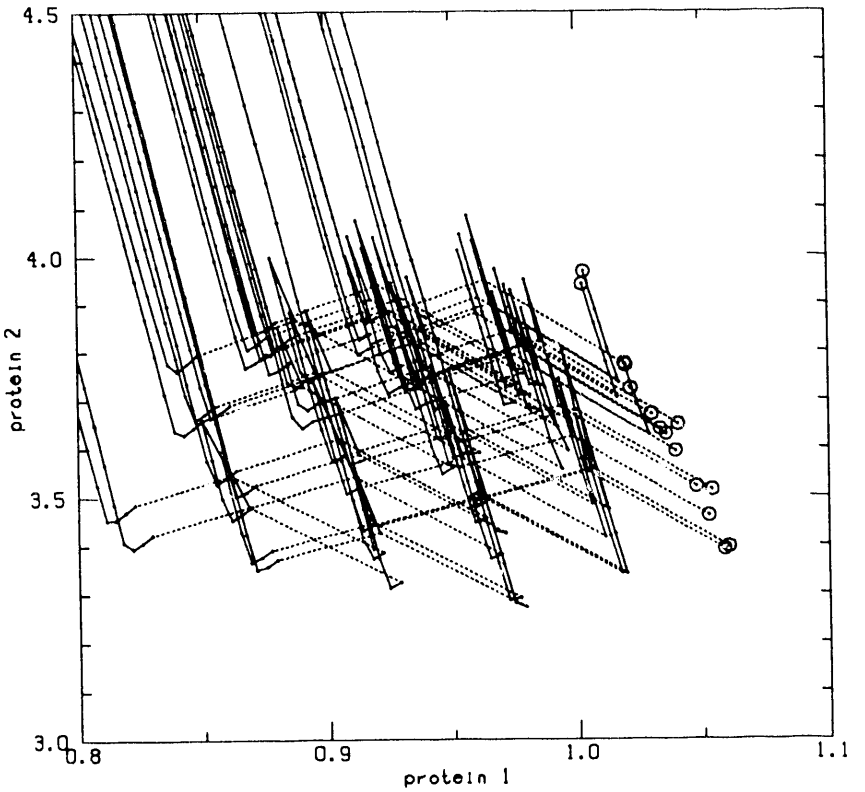

Figure 8: Phase portrait of the first exactly reproducing genotype, as expressed in slot seven. Note that this two-dimensional slice of a five-dimensional circuit state space reveals six groups of cells, which we may number from left (group 1) to right (group 6). Also, note that group 5 contains no terminally differentiated cells, groups 1 and 6 do not mitose, and mitosing cells move from group n to group $n \pm 1$.

[9] N. Metropolis, A. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[10] Eric Mjolsness, David H. Sharp, and John Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, 152:429–453, 1991.

[11] Przemyslaw Prusinkiewicz, Mark S. Hammel, and Eric Mjolsness. Animation of plant development. In *SIGGRAPH 93 Conference Proceedings*. Association for Computing Machinery Press, July 1993.

[12] John Reinitz, Eric Mjolsness, and David H. Sharp. Cooperative control of positional information in *Drosophila* by *bicoid* and maternal *hunchback*. Submitted, 1993.

[13] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.

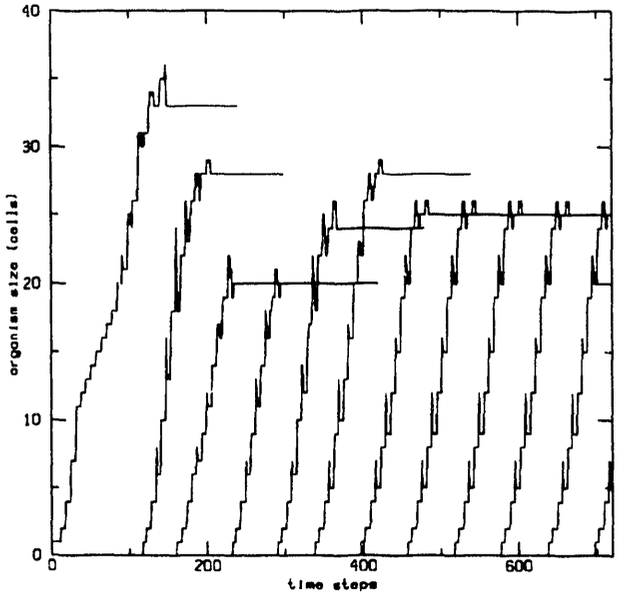

Figure 9: First exactly reproducing genotype, showing the size of the organism in each slot in the training set (the scoring function) as a function of environmental time. Note transient, followed by convergence to periodic behavior.

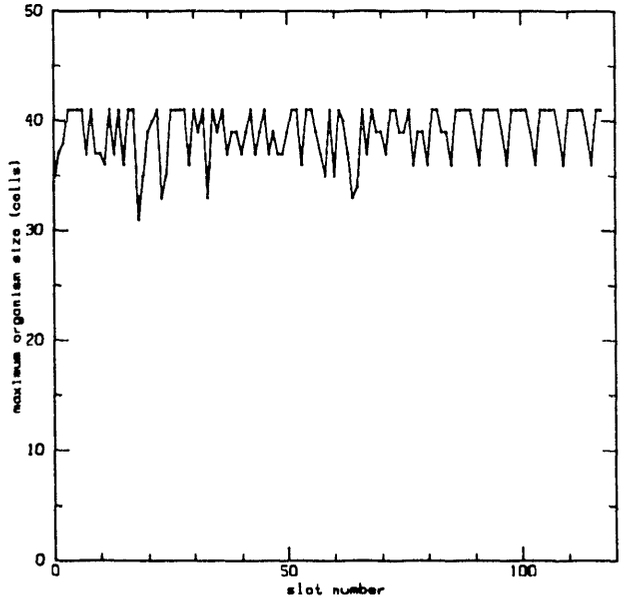

Figure 10: A period six life cyle. The maximum size of each slot's occupying organism is shown as a function of slot number (proportional to environmental time), for many slots. Only the first 12 slots were trained. Closer inspection of the organism-level lineage tree shows that this is actually three interleaved life cycles of period two.

DATE
FILMED
11/29/93
END