LA-UR 90-3923

CONF-9006299--1

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

LA-UR--90-3923

DE91 004850

TITLE: A Connectionist Model of the Drosophila

AUTHOR(S): John Reinitz, Columbia University
Eric Mjolsness, Yale University
David H. Sharp, Los Alamos National Laboratory

SUBMITTED TO: Proceedings of the Workshop on
Principles of Organization in Organisms
Santa Fe Institute, June 12-16, 1990

DISCLAIMER

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

A Connectionist Model of the *Drosophila* Blastoderm

John Reinitz*, Eric Mjolsness†[1], and David H. Sharp‡[2]

**Department of Biological Sciences, Columbia University*
New York, NY 10027

†*Department of Computer Science, Yale University*
P.O. Box 2158 Yale Station, New Haven CT 06520-2158

‡*Theoretical Division, Los Alamos National Laboratory*
Los Alamos, NM 87545

November 1, 1990

Abstract

We present a phenomenological modeling framework for development, and apply it to the network of segmentation genes operating in the blastoderm of *Drosophila*. Our purpose is to provide a systematic method for discovering and expressing correlations

[1]Supported in part by the Air Force Office of Scientific Research under grant AFOSR 88-0240.
[2]Work supported by the United States Department of Energy.

in experimental data on gene expression and other developmental processes. The modeling framework is based on a connectionist or "neural net" dynamics for biochemical regulators, coupled to "grammatical rules" which describe certain features of the birth, growth, and death of cells, synapses and other biological entities. We present preliminary numerical results regarding regulatory interactions between the genes *Kruppel* and *knirps* that demonstrate the potential utility of the model.

I. Introduction

We sketch a modeling framework for development. Its purpose is to provide a systematic method for discovering and expressing correlations in experimental data on gene expression and other developmental processes. In this report, we present preliminary results on the application of this modeling framework to the blastoderm of *Drosophila.* A further discussion of the underlying ideas is given in [11].

II. Basic Approach

The *configuration* of a developing embryo is specified by the number and internal state of its constituents. The constituents of an embryo include cells, cell nuclei, fibers, and synapses.

The choice of state variables depends on the object to be described. For example, we describe the state of of a cell nucleus in terms of concentrations of transcription factors, whereas a synapse could be described with membrane voltage, internal Ca^{++} concentration and ionic conductance.

Developmental dynamics is modeled by a set of equations describing the transition between two configurations (state history). The model describes three fundamental dynamical processes.

Concentrations of regulatory molecules acting within and between cells and their organelles change in response to three factors: existing concentrations of the regulators, exchange of regulatory molecules among existing entities, and simple decay. With the exception of decay, which is represented by a simple exponential decay term, these processes are modeled with "connectionist" or "neural net" dynamical equations, which incorporate basic phenomenological features of the regulatory molecules. At almost all times, the developmental dynamics is represented by a system of coupled ordinary non-linear differential equations, but when the number or kind of fundamental entities changes (for example when a cell divides) the system takes a discontinuous jump.

The continuous or discontinuous evolution of the system at a given moment of time is

modeled by a set of "grammatical rules" (see for example [10].). Rules are selected on an object to object basis, at each time. The rule selected by a particular object is determined by the state of that object, and of its neighbors. Each continuous dynamical process has an associated grammatical rule, in which there is no change in the type or number of fundamental entities. Changes in number and type of fundamental entities resulting from birth, growth, induction, and death processes are represented by grammatical rules that describe discontinuous dynamic processes. Finally, the model must describe the influence of spatial organization.

The combined action of continuous and discrete time grammatical rules is illustrated in Figure 1. The figure shows a representative history of state variables v_i^a under a combination of continuous and discrete time neural net dynamics. v_i^a is the concentration of regulator a in object i. Time increases to the right; three generations of objects are shown. Continuous time dynamics is denoted by the stippled horizontal axes; discrete time dynamics by the stippled ovals. The solid black curves are graphs of the functions $v_i^a(t)$. The rightmost oval denotes a type change; the other two are mitoses. The level of a state variable may change under the action of a discrete time rule. Note, for example, that $v^1(t)$ is always above baseline before a mitosis and and below it afterwards.

There are different classes of rules to represent the different classes of events that occur during development. In general, a biological object such as a cell or synapse may undergo a variety of transformations. It may be born from a parent or it may die. Between birth and death its internal state will change as a result both of internal dynamics and by interactions with other objects of the same or different type. Birth and death processes are represented by discrete time rules only; changes in internal state and interactions with other objects may be represented by either continuous or discrete time grammar rules. These possibilities taken together amount to six classes of rules.

Figure 2 is a diagrammatic representation of the six classes of rules which we employ. In each diagram, the time axis runs in a horizontal direction, and a space axis in the vertical direction. The arrowheads on the solid lines point in the direction of increasing time. The dotted vertical arrows represent a spatial interaction: continuous time rules by three such arrows, discrete time by one. The dotted arrows point in the direction of the object which has chosen the illustrated rule. A discrete time rule is represented by a filled circle; a continuous time rule by a pair of arrowheads on a solid line without an intervening filled circle. The input and output object types are indicated on the left and right hand sides respectively of each diagram. In (a), more than two branches could occur; the branches have lineage indices

i_n ranging from 0 to b.

The next two sections describe the application of this modeling framework to the *Drosophila* blastoderm.

III. Application to the *Drosophila* Blastoderm

Two features of the *Drosophila* blastoderm make it an especially suitable system for the initial application of the modeling framework outlined above. The first is that the regulatory dynamics of genes that lay out the basic body plan are dynamically separable from other aspects of the developmental dynamics during the blastoderm stage. The second is the availability of molecular probes for the products of these genes, which render the state variables v_i^a directly observable.

We review some elementary facts about the *Drosophila* blastoderm. Immediately following fertilization, the zygotic nuclei undergo a rapid series of mitoses without the formation of cells. After eight almost synchronous divisions, these nuclei migrate to the cortex of the egg, whereupon transcription of the zygotic genes begins. This stage is called the syncytial blastoderm, because no cells are present. After another five divisions, cell membranes are layed down and gastrulation begins [4]. The timing of these cell divisions is under the control

of maternal gene products. The protein products of zygotic pattern formation genes essential for laying down the basic body plan of the animal are expressed at this time in patterns that rapidly evolve from coarse to fine scale spatial resolution (reviewed in [1] and [7]).

The *Drosophila* egg is approximately an ellipsoid, but asymmetries in its shape clearly define two axes, each with a polarity. These axes provide coordinates for the blastoderm as well. One axis runs in an anterior-posterior direction, and the other in a dorsal-ventral direction. The pattern formation genes fall into two classes. To a reasonable degree of approximation, the level of expression of a member of the first class of genes is solely a function of location on the anterior-posterior axis; these genes are members of the anterior-posterior class. The expression level of of a member of the second class of genes depends only on position along the dorsal-ventral axis; these genes belong to the dorsal-ventral class.

The separation of these two classes of genes by expression pattern carries over to their dynamical interactions. A member of one of these classes of genes does not regulate the expression of a member of the other class during the blastoderm stage, except perhaps in the region of the anterior or posterior pole. For the rest of this report we focus on the zygotic anterior-posterior pattern formation genes, often referred to as segmentation genes. The segmentation genes are dynamically coupled in a network of genetic regulation. A line

of evidence leading to this conclusion is the observation that disabling one segmentation gene by mutation causes changes in the pattern of expression of many of the other segmentation genes [3, 8, 14, 5, 2, 12, 13]. The characterization of this regulatory network is one of the objectives of our modeling effort. A precise formulation of the regulatory network is required to interpret altered patterns of gene expression in terms of regulatory action.

The expression of the segmentation genes in the middle region of the blastoderm is approximately a function of anterior-posterior position only, so we model their dynamics of expression using a linear array of nuclei. At each time, a nucleus undergoes one of two types of transitions:

1. Mitosis: Replace each nucleus with a pair of daughter nuclei. Don't allow the synthesis of gene products. The mitoses are timed according to [4].

2. Interphase: Allow protein concentrations to evolve by synthesis, exchange of material with neighboring nuclei, and by degradation. This process is described by the dynamical equations for interphase.

We consider N genes. During interphase the level of gene product a in nucleus i, denoted

by v_i^a, is modeled by the the following connectionist equation (see [11]):

$$\frac{dv_i^a}{dt} = R_a g_a\Big(\sum_{b=1}^{N} T^{ab} v_i^b + h^a\Big) + D(n)\Big[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)\Big] - \lambda_a v_i^a. \qquad (1)$$

The first term describes gene regulation and protein synthesis, the second term describes exchange of chemical products between neighboring nuclei, and the third term describes the decay of gene products.

In (1), g_a is a thresholding function of the form $g_a(u^a) = (1/2)\Big((u^2/\sqrt{u^2+1}) + 1\Big)$ for all a, where $u^a = \sum_{b=1}^{N} T^{ab} v_i^b + h^a$. R_a is the maximum rate of synthesis from gene a. T^{ab} is a real number describing the influence of gene b on gene a. h^a is a threshold that is currently fixed at the same value of -10 for all genes. $D(n)$ depends on the number n of cell divisions that have taken place, such that $D(n) = 4D(n-1)$. λ_a is the decay rate of the product of gene a; the results given here assume a value of λ equivalent to a half life of 30 minutes. The other parameters T^{ab}, R_a, and $D(0)$ are adjustable to fit data.

IV. Preliminary Numerical Results

The question we investigate first is the regulation of the central domain of *knirps* expression by other zygotic segmentation genes. Among the genes known to regulate *knirps* are *Kruppel*,

hunchback, *giant*, and *tailless*. It is also known that *knirps* expression is reduced throughout the central domain in embryos mutant for *Kruppel* [12]. As a preliminary step, we have modeled the interaction of *Kruppel* and *knirps*. This truncated model may be valid in the posterior half of the *knirps* domain where *Kruppel* is believed to be the major regulator. This domain is approximately eight nuclei in extent along the anterior-posterior axis at the end of the blastoderm stage.

For comparison of the model with data, we rely on double labeling studies using fluorescence tagged antibodies (unpublished data of R. Warrior and J. Reinitz). Some embryos were photographed under bright field optics as well as fluorescence in order to more accurately assess the developmental stage. Our model requires specification of initial conditions: these are illustrated in Figure 3, which shows the distributions of *Kruppel* and *knirps* proteins at the beginning of cleavage cycle 13. Given these initial conditions, we model the dynamics of gene products from that time until the onset of gastrulation, a period of about 70 minutes. During that period one nuclear division takes place and the number of nuclei doubles.

To discover what values of T^{ab}, $D(0)$, and R_a best fit the data, we do a least-squares fit to the *trajectories* that the system follows. Given a set of fixed initial conditions (e.g. Figure 3), each v_i^a will follow a trajectory that depends on the values of the parameters in (1). Our

aim is to find values of those parameters such that the trajectories given by (1) are as close as possible to the observed trajectories of changing protein concentrations in each cell. As an approximation to the observed trajectory, we compare the model to gene expression data at two times: mid cleavage cycle 14 and the onset of gastrulation. For example, the expression pattern just prior to gastrulation is shown in Figure 4. We measure the deviation between the the data and the behavior of the model by taking the sum of squared differences between the observed protein concentration and that given by the model over each protein, cell, and time for which data exists. This deviation is then minimized using the method of simulated annealing [9].

The behavior of the model after such a fit is shown in Figure 5. Results obtained using wild-type data allow us to predict the behavior of mutants without further experimental input. In particular, we find *knirps* expression greatly reduced in mutants for *Kruppel*; *Kruppel* expression extends more posteriorly in mutants for *knirps*. This is in qualitative agreement with experimental observations [6]. Although the numerical results reported here are quite preliminary, they indicate that the methods we use promise to be helpful in characterizing the network of genes that control pattern formation in *Drosophila* and other organisms. We are currently investigating a more comprehensive model of the blastoderm which includes

the effect of other genes.

References

[1] Michael Akam. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development*, 101:1–22, 1987.

[2] Sean B. Carroll, Allen Laughon, and Bruce S.Thalley. Expression, function, and regulation of the *hairy* segmentation protein in the *Drosophila* embryo. *Genes and Development*, 2:883–890, 1988.

[3] Sean B. Carroll and Matthew P. Scott. Zygotically active genes that affect the spatial expression of the *fushi tarazu* segmentation gene during early *Drosophila* embryogenesis. *Cell*, 45:113–126, 1986.

[4] Victoria A. Foe and Bruce M. Alberts. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *Journal of Cell Science*, 61:31–70, 1983.

[5] Manfred Frasch and Michael Levine. Complementary patterns of *even-skipped* and *fushi tarazu* expression involve their differential regulation by a common set of segmentation genes in *Drosophila*. *Genes and Development*, 1:981–995, 1987.

[6] K. Harding and M. Levine. Gap genes define the limits of antennapedia and bithorax gene expression during early development in *Drosophila*. *The EMBO Journal*, 7:205–214, 1988.

[7] P. W. Ingham. The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature*, 335:25–34, 1988.

[8] H. Jackle, D. Tautz, R. Schuh, E. Seifert, and R. Lehmann. Cross-regulatory interactions among the gap genes of *Drosophila*. *Nature*, 324:668–670, 1986.

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

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

[11] Eric Mjolsness, David H. Sharp, and John Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, Submitted, 1990.

[12] M. J. Pankratz, M. Hoch, E. Seifert, and H. Jackle. *Kruppel* requirement for *knirps* enhancement reflects overlapping gap gene activities in the *Drosophila* embryo. *Nature*,

341:337–340, 1989.

[13] John Reinitz and Michael Levine. Control of the initiation of homeotic gene expression by the gap genes *giant* and *tailless* in *Drosophila*. *Developmental Biology*, 140:57–72, 1990.

[14] Christine Rushlow, Katharine Harding, and Michael Levine. Hierarchical interactions among pattern-forming genes in *Drosophila*. In *Banbury Report 26: Developmental Toxicity: Mechanisms and Risk*. Cold Spring Harbor Press, 1987.

Figure Legends

Figure 1

A schematic illustration of the history of the state variables v_i^a under a combination of continuous and discrete time grammatical rules.

Figure 2

A diagrammatic representation of the six classes of grammatical rules.

Figure 3

Distribution of the products of the genes *Kruppel* and *knirps* over a domain containing 4 nuclei at the beginning of cleavage cycle 13. The vertical axis is calibrated in arbitrary units of protein concentration; its scale was selected to be same as that used in Figures 4 and 5. The horizontal axis is calibrated in terms of individual nuclei in a line running along the anterior-posterior axis; nucleus 1 is the most anterior. Filled circles denote *knirps* concentrations; open circles denote *Kruppel* concentrations. Note that the only non-zero concentration is that of *Kruppel* in nucleus 1.

Figure 4

The graph shows our estimate of the relative levels of expression of *Kruppel* and *knirps* in a strip of eight nuclei running from the middle of the central *knirps* domain in a posterior direction. This estimate was made from photographs of embryos stained with the appropriate antisera (see text). The vertical axis is calibrated in arbitrary units of protein concentration. The horizontal axis is calibrated in terms of individual nuclei in a line running along the anterior-posterior axis; nucleus 1 is the most anterior. A cell division has occurred since the situation shown in Figure 3, so that nuclei 1 and 2 are daughters of nucleus 1 in figure 3, and so on. Filled circles connected with dotted lines denote *knirps* concentrations; open circles connected with solid lines denote *Kruppel* concentrations.

Figure 5

The graph shows the numerical output of the model for wild type and one mutant at the onset of gastrulation. The *Kruppel-Kruppel* connection strength was 5.1, the *Kruppel-knirps* connection strength was 3.9. The two corresponding connection strengths from *knirps* to *Kruppel* and itself were –.35 and 1.2 respectively. $R_{Kruppel}$ was .83, R_{knirps} was 3.4, and $D(0)$ was .008. The axes and symbols are as described for Figure 4, with the addition that the

open squares connected by dashed lines represent the distribution of *Kruppel* in an embryo numerically mutated for *knirps*; note that it extends more posteriorly than in wild type, in accordance with observations. Expression of *knirps* in this numerical mutant and of both *Kruppel* and *knirps* in a numerical mutant for *Kruppel* are not shown; all three were very close to zero in all nuclei.

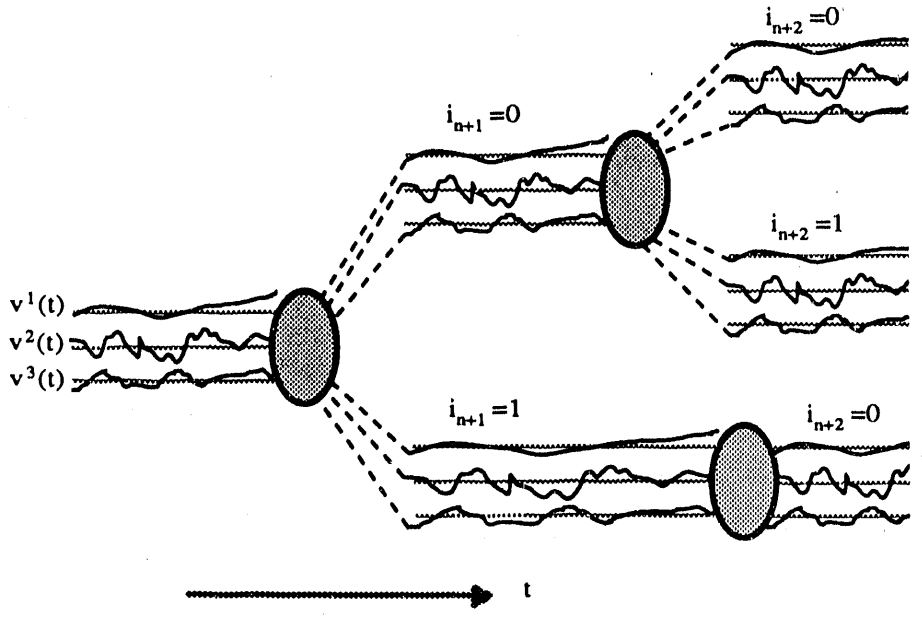
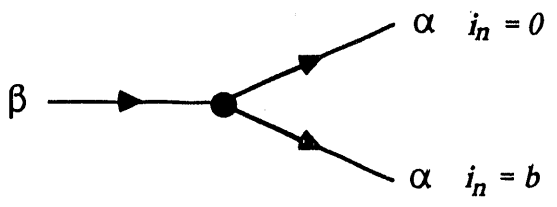


Fig. 1

Discrete Time Rules:

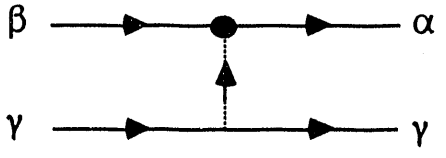
(a) Birth:



(b) Death:

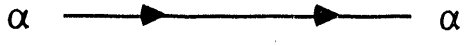
(c) One Object:

(d)Two Objects:



Continuous Time Rules:

(e) One Object:



(f)Two Objects:

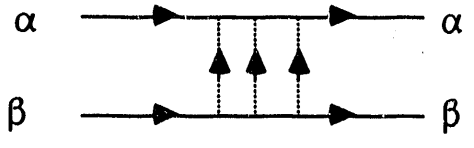


Fig. 2

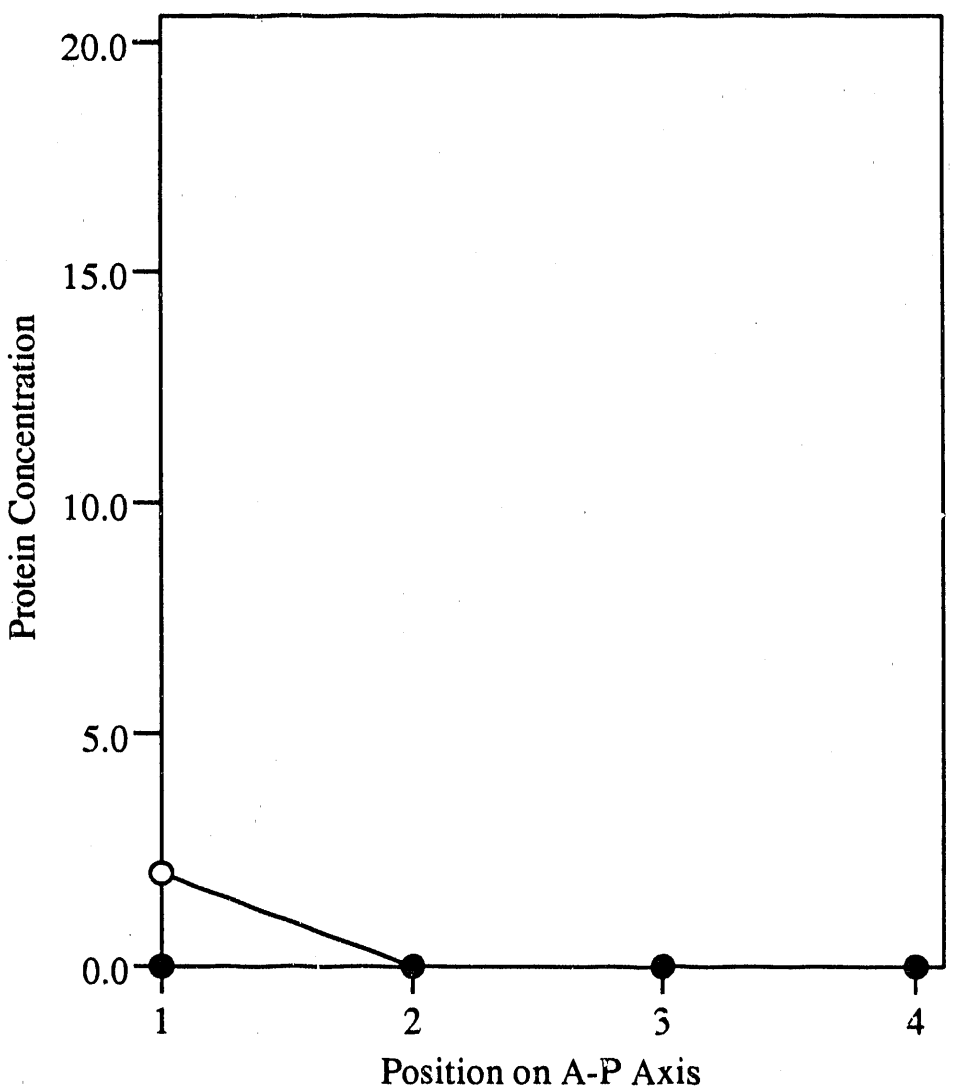

20.0
15.0
10.0
5.0
0.0
Protein Concentration
1
2
3
4
Position on A-P Axis

Fig. 3

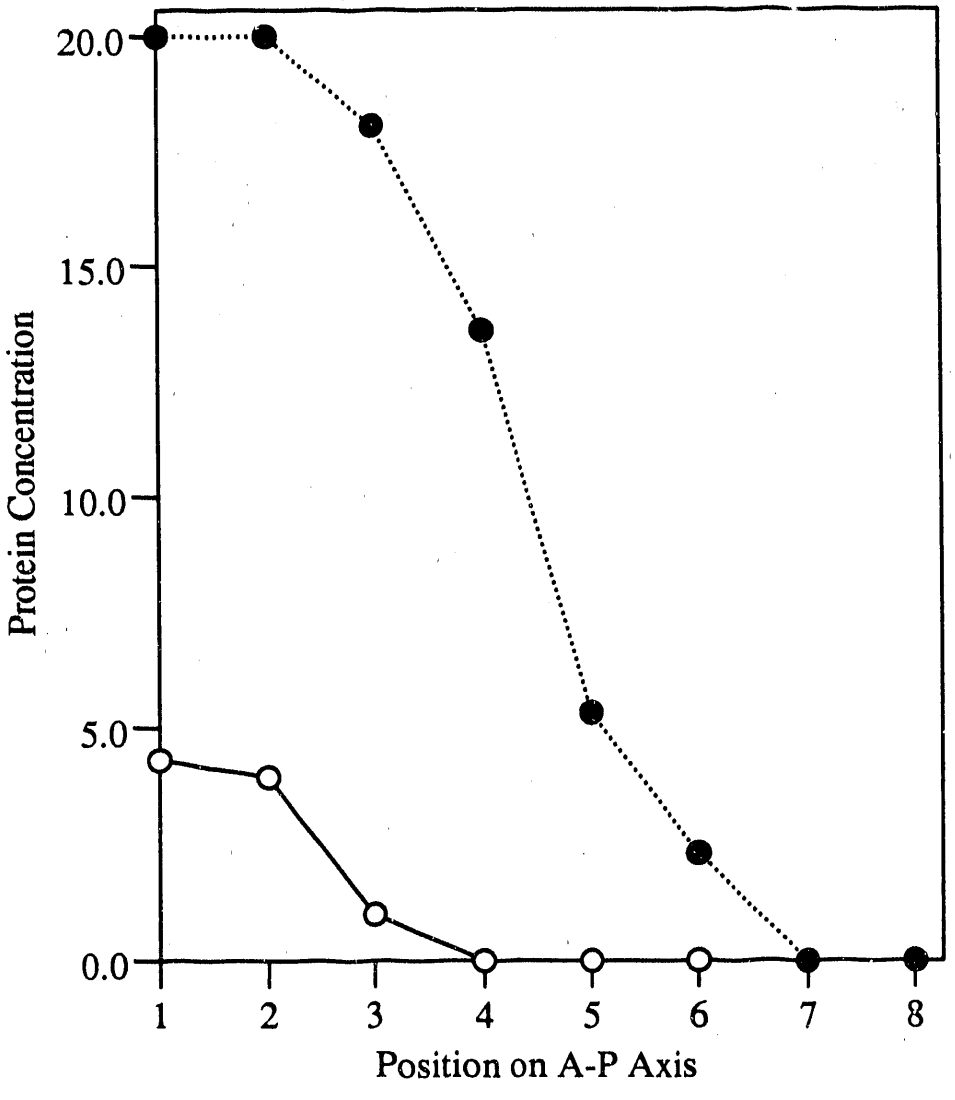

20.0
15.0
10.0
5.0
0.0
Protein Concentration
1
2
3
4
5
6
7
8
Position on A-P Axis

Fig. 4

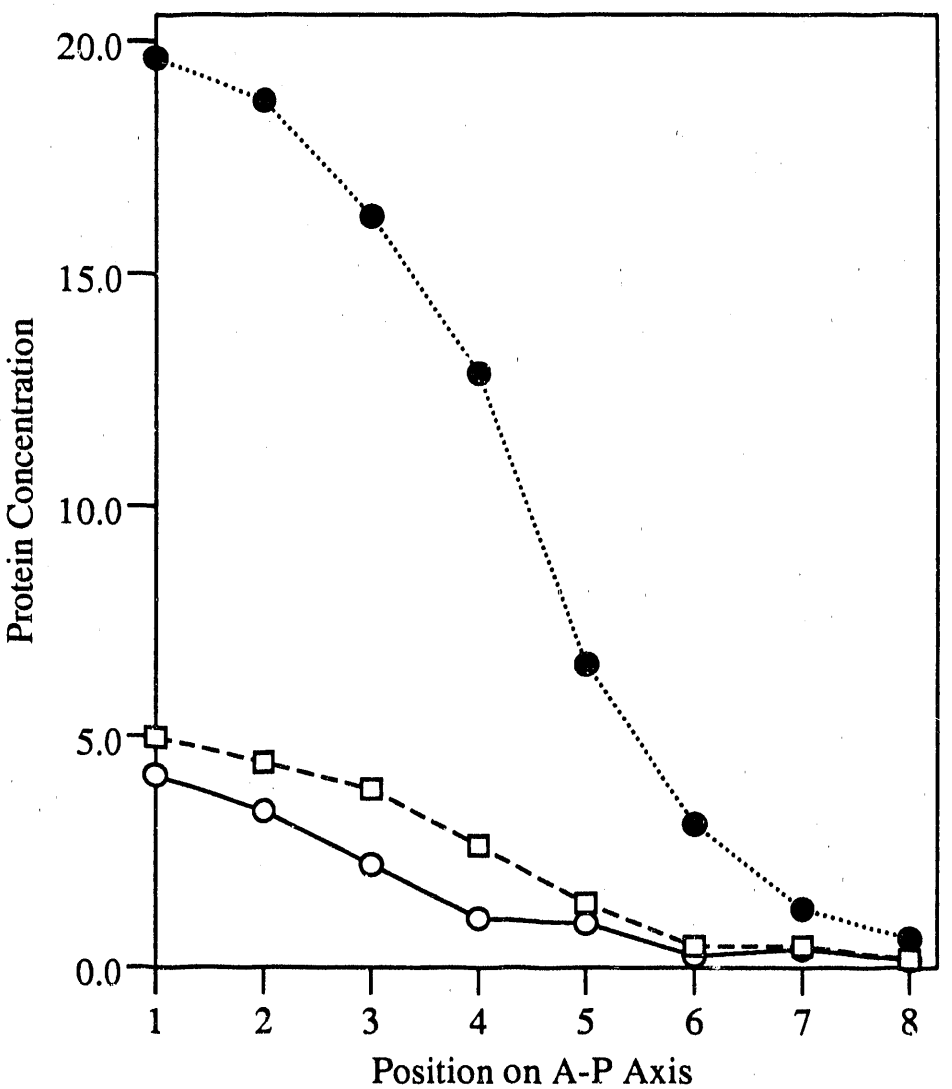

20.0
15.0
10.0
5.0
0.0
Protein Concentration
1
2
3
4
5
6
7
8
Position on A-P Axis

Fig. 5

END

DATE FILMED

02/27/91