

Boolean modeling of gene regulatory networks: Driesch redux

David N. Arnosti^{a,1} and Ahmet Ay^{b,1}

^aDepartment of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824; and ^bDepartment of Mathematics, Department of Biology, Colgate University, Hamilton, NY 13346

Conducting research on sea urchins at the Naples Zoological Station, 19th century developmental biologist Hans Driesch demonstrated the totipotent nature of early embryonic cells, contributing significantly to the then-nascent field of “developmental mechanics.” Driesch discussed the possibility of understanding the clocklike development of this organism in physical/mathematical terms, but ultimately retreated from this stand later in his career (1). As developmental biology moves into its third century of existence as a modern science, we find that major advances are bringing us full-circle to approach central questions posed by early pioneers (Fig. 1). In PNAS, Peter et al. (2) describe a quantitative model to describe at a molecular level the processes of cellular differentiation that have fascinated generations of biologists, providing a means to link developmental and systems biology.

During the past four decades, this group of researchers (2) has exploited the sea urchin *Strongylocentrotus purpuratus* embryo to develop the notion of the “hard-wired” gene regulatory network (GRN), in which cascades of regulatory events serve to differentiate specific lineages of cells in the embryo. These regulatory steps involve activation and repression of gene expression by transcriptional enhancers, which integrate and transduce the activity of cell-specific factors. With its readily traceable cell lineages, along with appropriate technologies for visualizing and manipulating gene expression, the sea urchin embryo provides a powerful system in which to identify key elements of regulatory circuits. Discoveries during the past two decades have confirmed that key signaling systems are highly conserved in metazoan development; thus, from the standpoint of genetic circuitry involved, the sea urchin embryo serves as a general model for animal embryogenesis (3).

To identify the components of the GRN, many years of extensive experimental analysis have concentrated on identifying *cis*-regulatory elements and specific transcription factors of developmental genes. The current GRN for endoderm and mesoderm includes ~50 regulatory genes active from early cleavage stages (6 h after fertilization) to the onset of gastrulation (30 h after fertilization). The results of such studies, laid out in a circuit-board fashion, are impressive, but can prompt feelings of confusion akin to those induced by

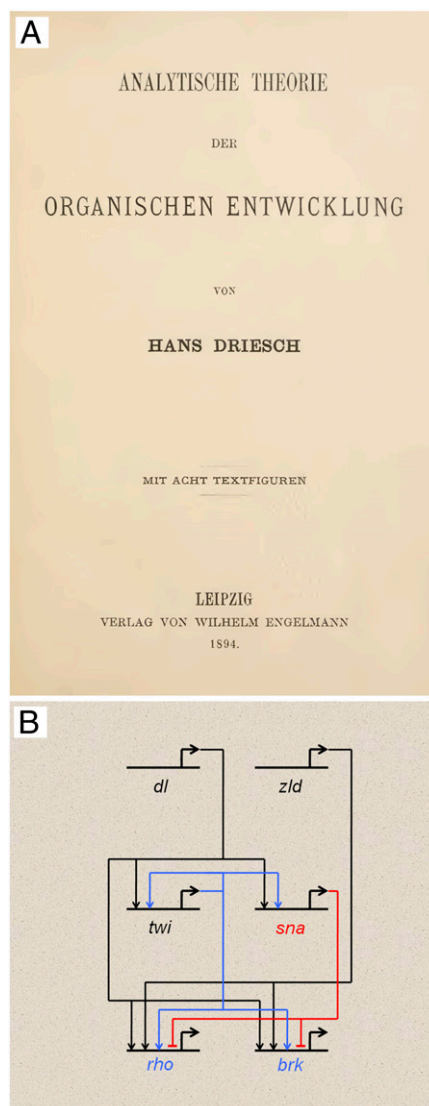


Fig. 1. (A) Title plate for developmental biologist Hans Driesch's *Analytical Theory of Organic Development* (1894), in which the author summarizes the mechanistic understanding of embryonic development at the time. (B) An example of an embryonic GRN summarizing three-layer molecular interactions among transcriptional regulators specifying primitive mesoderm and neurectoderm of *Drosophila*. The complex interactions among regulatory genes can be modeled by using Boolean operators to simulate the tissue specification observed in vivo.

the familiar protein–protein interaction “hairball” diagrams from systems biology studies. What is the utility of such representations to biologists? Some features of

these diagrams are apparent by inspection; many gene regulatory circuits feature feedback and feedforward circuit design, which theoretical studies indicate can enhance the precision and robustness of regulation (4). Analysis of GRNs should also allow us to understand how variation—whether environmental or genetic—might lead to alternative operation of the GRN. Well-documented studies indicate that such regulatory divergences play important roles in development, disease, and evolution (5).

Mathematical modeling of gene expression can provide insights into the activity and structure of individual genes or entire gene circuits. Dynamic representations of gene expression have been approached by using differential equations, whereby specific terms allow one to represent the spatially and temporally changing patterns of gene expression. These models are, in general, computationally more demanding than simpler Boolean models, in which logic statements with simple “and”/“or”/“not” statements represent complex biochemical processing that occurs in signaling and transcriptional switches. Such statements can be coupled into chains to generate a temporal sequence of events involving gene activation and inactivation, with time steps selected by the investigator (6). By using the pair-rule gene expression network of the *Drosophila* embryo as a test case, Albert et al. have shown that Boolean representations of GRNs can achieve similar levels of accuracy to more complex ordinary differential equation models (7, 8).

Modeling can be used for investigative purposes, especially for systems with poorly described parameters. For such systems, a variety of different network configurations can be analyzed to obtain insights into the fundamental nature of the underlying system (7–9). In contrast, for systems with detailed information, such as the one studied in the work of Peter et al. (2), modeling can also be used for summarizing the system and validating the knowns. A previous study from the same laboratory provided one of the first examples of Boolean modeling of eukaryotic transcriptional switch, summarizing and validating experimental information for *endo16* promoter region (10). In

Author contributions: D.N.A. and A.A. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 16434 of issue 41 in volume 109.

¹To whom correspondence may be addressed. E-mail: arnosti@msu.edu or aay@colgate.edu.

that study, detailed interactions of different sets of *cis*-regulatory elements of the *endo16* promoter were represented by a set of logic statements, and the model demonstrated that these binary representations of transcriptional regulation were adequate to reproduce the spatial and temporal regulation of this endoderm marker. This modeling effort provided a synthesis of information about a single gene, but did not generalize to whole circuits or other genes.

In the study of Peter et al. (2), the authors apply Boolean modeling to represent the gene regulatory events in four primary tissues of the early sea urchin embryo. Similar to the Boolean analysis of *endo16*, the authors write down logic statements that are derived from extensive experimental measurements. However, this project is on a much larger scale than any attempted before. Being tied to the specific regulatory interactions measured in *S. purpuratus*, this model is not necessarily directly applicable to the understanding of other embryos or other GRNs. Rather, the objectives here are to understand whether the current complex (but by no means complete) GRN is sufficiently informative to predict the major transitions in early development, and whether the simple representations of the GRN capture the most important elements of the developmental program.

The model examines gene expression in two domains of the mesoderm (skeletal and nonskeletal), and the future anterior and posterior endoderm domains over the course of 1 d, from early cleavage stages to the start of gastrulation. By using a time step of 3 h, nearly 3,000 time/space calculations were performed for the almost 50 regulatory genes. There were some deviations observed between the model and experimental observation, including the late expression of *foxa*, the inputs of which are unknown. Reminiscent of a high school band practice, most of the deviations generated by this model were of temporal nature (i.e., late entrances or playing through the rests), whereas only one spatial discrepancy was observed. The small percentage of such deviations indicate the overall completeness of this modeling scheme at the time step resolution selected. Interestingly, extending the intervals from 3 h to 4 h significantly decreased model performance, indicating that, for this stately developmental un-

folding of gene expression, there are specific time windows that capture critical events.

A significant test of this model was the prediction the effects of specific perturbations of the overall GRN. The knock-down of Delta signaling molecule or overexpression of the Pmar repressor have dramatic effects *in vivo*, and can be simulated by manually turning down or activating these genes in the model. The known downstream effects were indeed observed, as were extensive changes in other components that have not been studied experimentally. More dramatically, a classical transplantation experiment by Hörstadius is modeled here. Such embryo surgeries, used most famously by Spemann and colleagues, showed that specific cells in the developing animal embryo express inducing substances that have the capacity to transform developmental fates of neighboring cells. Here, cells from the sea urchin vegetal pole were transplanted to the animal pole, inducing gut-specific differentiation programs typical of the vegetal axis. The Boolean model can capture the initial stages of gene expression from such an intervention.

The success of their Boolean model persuades the authors that the directed developmental changes of the sea urchin embryo are indeed Boolean in nature (2); as they point out, downstream genes appear to respond to input modules well before the latter have reached a steady state, indicating that it is not the fine gradations in output that are critical. Other studies from morphogen-like factors such as TGF- β family members indicate that differentiation signals are exquisitely sensitive to signaling levels; thus, the success of the rough Boolean representation may actually reflect the robust circuitry that is built into differentiation pathways. Furthermore, their model does not include regulation by microRNAs, which play pervasive functions in eukaryotic gene regulation, although frequently only in a supportive “shock-absorber” role. Normally robust regulatory pathways exhibit self-correcting behavior, which begins to break down when important nodes or connections are weakened or disabled (11, 12).

As the authors note (2), their approach is deterministic; given specific starting conditions (maternal inputs of transcrip-

tion factors and signaling molecules), the stepwise unfolding of gene transitions will be reproducible. However, even robust GRNs typically show considerable variation in activity at the molecular level. It will be interesting to learn how such Boolean models function when married to stochastic operators that qualify the output of the logic circuits with probability statements.

This study (2) raises important questions to be addressed in future studies. The broad outlines of the embryonic GRN are represented in a simplified fashion, with arrows indicating inputs from a signaling pathway to specific genes. These inputs actually involve proteins binding to complex DNA regulatory elements, and we know there is tremendous variation in *cis*-regulatory elements even on a population level. For example, the sequence of the paradigmatic *endo16* promoter was found to be dramatically different in individuals collected from the same coastal area; does this variation reflect selective pressure for alternative outputs, or are dramatically different promoters roughly equivalent in overall function (13)?

Looking ahead, what are the most promising objectives for systems-wide modeling of GRNs? Current technology is poised to deliver massive amounts of personalized genome information in the field of medicine, and we will soon be able to access with ease genomic information at a population and species level on any system. Surely one of the most important aims in gene regulatory modeling must be to understand how DNA variation, encoded in genomic sequences, translates into distinct GRN outputs. The recently released ENCODE studies point to the torrents of data on genome structure, gene expression, and chromatin now available to the modern biologist, a treasure trove that Hans Driesch could not have dreamt of. However, we still are challenged by the complexity of developmental processes—how can we understand them in physical/mathematical terms? Clearly, mathematical models can help us understand the (in)completeness of our understanding of essential processes in development, at which point we may indeed write a second edition of *Analytical Theory of Organic Development*.

1. Driesch H (1894) *Analytische Theorie der Organischen Entwicklung* (Engelmann, Leipzig, Germany).
2. Peter IS, Faure E, Davidson EH (2012) Predictive computation of genomic logic processing functions in embryonic development. *Proc Natl Acad Sci USA* 109:16434–16442.
3. Davidson EH (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic, San Diego).
4. Hart Y, Antebi YE, Mayo AE, Friedman N, Alon U (2012) Design principles of cell circuits with paradoxical components. *Proc Natl Acad Sci USA* 109:8346–8351.
5. Wunderlich Z, et al. (2012) Dissecting sources of quantitative gene expression pattern divergence between *Drosophila* species. *Mol Syst Biol* 8:604.

6. Ay A, Arnosti DN (2011) Mathematical modeling of gene expression: A guide for the perplexed biologist. *Crit Rev Biochem Mol Biol* 46:137–151.
7. Albert R, Othmer HG (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J Theor Biol* 223:1–18.
8. Chaves M, Albert R, Sontag ED (2005) Robustness and fragility of Boolean models for genetic regulatory networks. *J Theor Biol* 235:431–449.
9. von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust developmental module. *Nature* 406:188–192.
10. Yuh CH, Bolouri H, Davidson EH (2001) *Cis*-regulatory logic in the *endo16* gene: Switching from a specifica-

- tion to a differentiation mode of control. *Development* 128:617–629.
11. Jaeger J, et al. (2004) Dynamic control of positional information in the early *Drosophila* embryo. *Nature* 430:368–371.
12. Kozlov K, Surkova S, Myasnikova E, Reinitz J, Samsonova M (2012) Modeling of gap gene expression in *Drosophila kruppel* mutants. *PLOS Comput Biol* 8:e1002635.
13. Balhoff JP, Wray GA (2005) Evolutionary analysis of the well characterized *endo16* promoter reveals substantial variation within functional sites. *Proc Natl Acad Sci USA* 102:8591–8596.