Graph Grammars as Models for the Evolution of Developmental Pathways

Martin Beck[1], Gil Benkö[2,3], Gunther J. Eble [3,4] Christoph Flamm[5], Stefan Müller[5], Peter F. Stadler[1,3,5]
[1]Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig
[2]Graduiertenkolleg Wissensrepräsentation, Universität Leipzig
[3]Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig
Kreuzstrasse 7b, D-04103 Leipzig, Germany
[4] Centre National de la Recherche Scientifique, UMR 5561 Biogéosciences, Université de Bourgogne, Dijon 21000, France
[5] Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien
Währingerstrasse 17, A-1090 Wien, Austria
xtof@tbi.univie.ac.at

Abstract

The large quantity and ready availability of developmental-genetic data, coupled with increased rigor and detail in the characterization of morphological phenotypes, has made the genotype-phenotype map of whole organisms a central challenge in evolutionary developmental biology. This in turn necessitates more general modeling strategies that can efficiently represent different types of biological knowledge and systematically applied across levels of organization, spatiotemporal scales, and taxonomic groups. Graph-based models appear useful in this context but have been remarkably underutilized in biology. Simulation of ontogenetic and evolutionary change by means of graph-rewriting algorithms has been explored as a means of providing a coordinate-free approach to form transformation in time and space. A finite set of rules describing generic graph transformations is used to encode knowledge about morphogenetic steps. Their application to skeletal growth in sea urchins effectively models ontogenesis in terms of topology rather than specific geometry, suggesting a promising approach to general modeling of developmental evolution.

1 Introduction

Molecular evolution is firmly grounded in the Neo-Darwinian principle that all heritable variation is introduced into the next generation by means of mutation, recombination, or other genetic operators, while selection and other sorting processes differentially act on these variations at the level of the phenotype. Before the fate of a new phenotype can be determined, that phenotype must first be produced, or *accessed*, by means of variational mechanisms [8]. The genotype-phenotype map therefore takes center stage in any theoretical or computational attempt to model evolutionary changes [9, 17, 25, 26].

In the simplest case — evolving RNA molecules — genotype and phenotype are two aspects of the same molecule. The specific sequence of nucleotides is the genotype, the three-dimensional shape of the molecule represents its phenotype [23]. A series of computer simulations using RNA secondary structures as model phenotypes showed that phenomena such as neutral drift, punctuated change, plasticity, environmental and genetic canalization, and the emergence of modularity, can be reproduced within this framework, see e.g. [1, 9, 13, 23]. Concomitantly, specific predictions about RNA evolution models have been verified experimentally [22, 24]. Despite the success of the RNA model, and the fact that this approach can at least in part be extended to protein evolution [2, 16], it seems impossible to generalize detailed, biophysically accurate models (that derive all properties of an organism "*ab initio*" from the

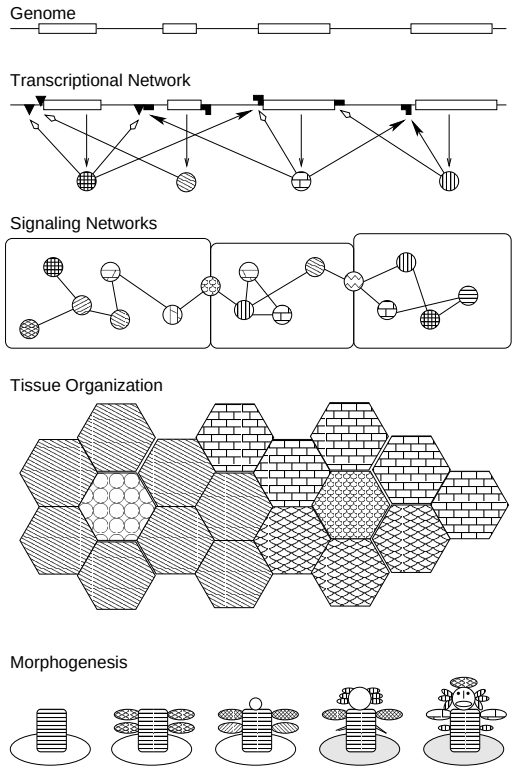

Figure 1: A comprehensive computational model that can be used to study the evolution of multicellular organisms must represent at least the main organizational levels known in plants and animals. The eventual outcome of the ontogenetic and epigenetic processing of information in genomic DNA sequences (top) is the (adult) phenotype (lower right), which may engage in fitness dynamics. As far as we know, even large-scale innovations of bodyplans map down to (originally) small changes at the genetic level that are amplified through the intermediate levels of gene expression, cell-cell signaling, and tissue determination and organization that eventually structure the ontogenesis (development) of the organism from a fertilized egg to the adult stage. From the modelling point of view, each level is best implemented separately, raising the issue of how different levels of description interact, and how this interaction itself can be modeled appropriately.

genomic sequences) much beyond extremely simplified unicellular organisms such as the one used to simulate the evolution of primitive genetic codes in [27].

Comprehensive computational models of more complex life forms, and in particular multicellular organism, thus have to reflect the multiple organizational levels of these systems, Fig. 1. Given the limitations of present-day computers, and the incompleteness of our knowledge about the structure and dynamics of each focal level as well as of interlevel interactions, we suspect that this hierarchy of organizational levels can be most fruitfully studied using higher-order descriptions specifically applied to each level. In essence, these descriptions then form phenomenological theories of different aspects of an organism. The important issue here is that at each level we ideally would like to include the possibility of evolvability, i.e., of the appearance of novel structures and functions that might well be beyond our current biological knowledge.

Most of today's models of gene regulatory networks [3, 7, 10, 15, 21], for example, are based on the well established "operon model" of gene regulation [14]. The operon model distinguishes between two types of genes, "regulatory" ones, encoding for transcription factors and "structural" genes, transcribed to proteins that play some functional role, e.g. in the cell's metabolism. Models based on this simple protein-based regulatory logic can show surprisingly intricate dynamical behaviours, ranging from complex periodic patterns to self-organisation and chaos, and thus would seem sufficient to understand this level of organization. The discovery of hundreds of microRNAs, see [20] for a review, thus came as a surprise. Eukaryotic cells use these tiny RNAs as an additional, and altogether different mode of gene regulation acting post-transcriptionally, presumably helping to overcome the complexity limitations arising in very large regulatory networks [4, 18]. In principle, innovations of this type might be investigated in large-scale computer simulations of evolution.

As one moves up the hierarchy of biological organization, emergent structures and functions codify additional complexity, and inter-level coordination becomes as prominent as phenomena at each focal level. Phenotypes become individuated, and from their integration whole organisms are consolidated.

In multicellular organisms, development can profoundly influence the possibilities of evolution. Indeed, modeling the development and evolution of complex phenotypes and of whole organisms has always been a major goal in theoretical biology.

Epigenetic models of morphogenesis as well as models of developmental gene expression and regulation have yielded important insights, but traditionally they have focused on particular taxa, morphological modules, and gene expression domains. As a result, their generality remains limited. In contrast, evolutionary developmental biology today is faced with an abundance of developmental-genetic data, on the one hand, and with more detailed and rigorous descriptions of the emergence of morphological characters and whole-organism form during ontogeny, on the other. With the realization that genotypes and phenotypes are manifested and can interact in multiple spatiotemporal contexts and across organizational levels, understanding the genotype-phenotype map as it relates to organismal development and evolution presents itself as a research agenda on the very fullness of biological complexity. Accordingly, there is a pressing need for novel modeling approaches that explicitly allow for multiple levels and for the representation of whole-organism integration, and that are more general by providing computational platforms flexible with respect to dimensionality and parameterization, and therefore applicable to a broad array of taxa, regardless of organismal geometry.

We are of course not claiming to meet this challenge *in toto*, but rather to underscore the need for a new class of models of developmental evolution. In this contribution we merely show how the computationally powerful framework of graph grammars can be employed to model one aspect of whole-organism ontogeny that has received relatively little attention by the artificial life community in the past: the construction of the adult phenotype from the juvenile phenotype in a strictly epigenetic dynamics of morphological units (which are also the relevant raw material for considerable macroevolutionary change).

2 Ontogenesis as Graph Rewriting

An "organism" is abstracted as a graph with adjacency representing spatial relationships and vertex labels and edge labels representing tissue types and interaction "classes". Rewrite rules from a graph grammar are used to transform the graph, Fig. 2. In general, there are two types of rewriting rules: (i) rules that change the connectivity of the graph and (ii) rules that only change the edge and/or vertex labels. The latter type of rule can be used to propagate signals along the graph scaffold.

We propose this computational implementation as an alternative to standard morphospace representations in the field of theoretical morphology [19] for two reasons: (i) innovation cannot be described properly in a setting in which a morphology is simply described as a point in a vector space and (ii) any attempt at detailed mechanistic modeling of tissue formation and growth in 3D space [6] would simply exceed our computational resources, at least for the purpose of large-scale simulations of evolution and the exploration of parameter space. This approach also differs from the empirical framework of geometric morphometrics (which allows for sophisticated descriptive models) in that the goal is neither to empirically represent spatially explicit transformations among neighbor morphologies, nor to consider them relative to constructs such as the mean shape, invariably influenced by the samples chosen. Rather, graphs are used as more general representations of the relational configuration and connectivity of skeletal elements. Therefore, our approach suggests coordinate-free protocols of collection of empirical data for the purposes of comparison with theoretical results.

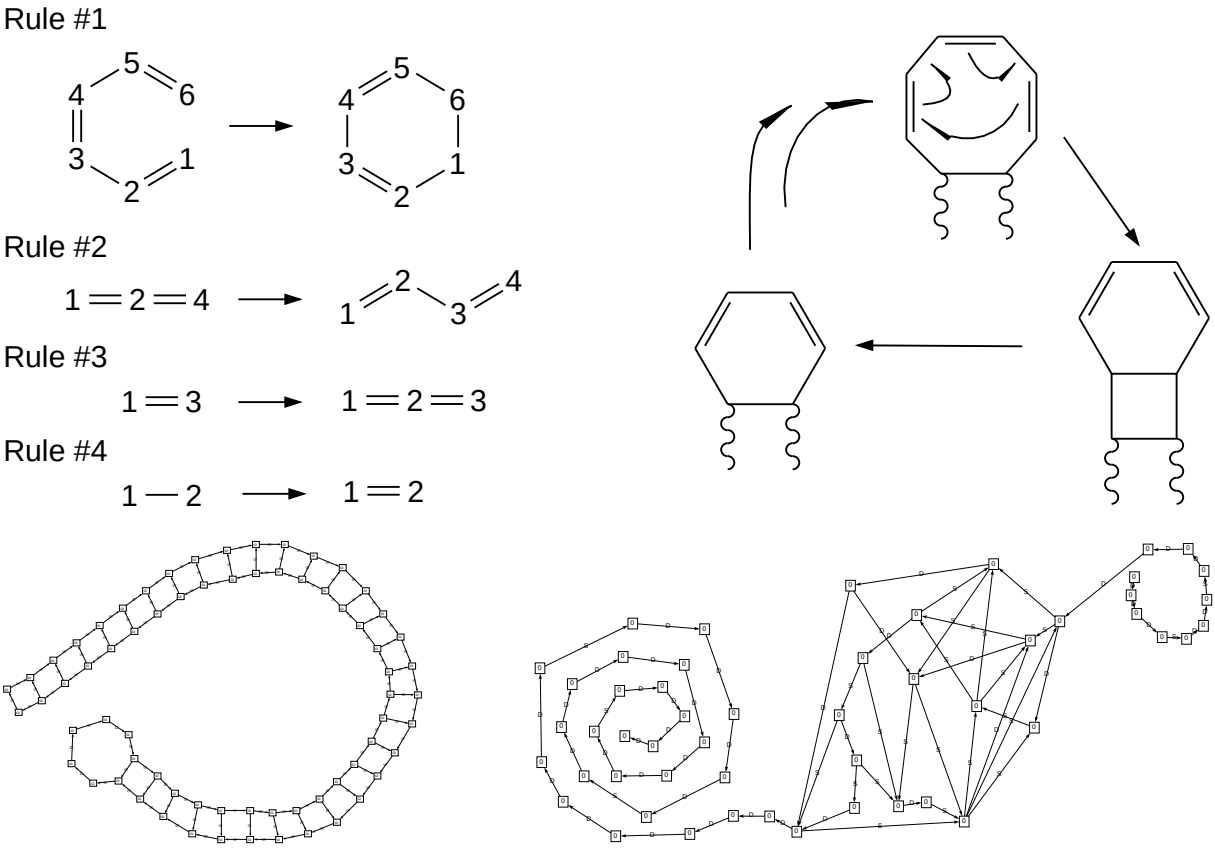

Figure 2: Graph rewriting as a minimal model for the evolution of development. The upper left panel gives the rewriting rules. The upper right panel depicts the growth cycle of an idealized "worm" when the rules are hierarchical, i.e. the first rule in the list that can be applied is applied to the graph. The result is shown in the lower left panel. The lower right panel gives the phenotype of a "developmental mutation" that destroys the order of the list of rules, i.e., the rules are applied in random order. Note that the level of organization and the units of construction need not be specified.

3 Implementation Issues

We decided to implement the graph rewrite framework as a distributed system consisting of several central servers each processing a specific set of rewrite rules and of many clients each hosting a "developing phenotype". A client starts a simulation cycle by sending a first graph to one of the central servers. The server tries to rewrite the graph by applying the first matching rule to the graph. The client then recieves either the transformed graph or the original one signaling that no further rule is applicable. Graphs are exchanged in the *graph modelling language* (GML) [11] between client and server. The algorithmic framework of the graph-rewrite engine is based on the class of UBS graph-rewriting systems [5], for which the complexity of a rewriting step is linear. The core of the graph-rewrite engine is written in the functional programming language `Haskell` [12].

4 First Simulation Results

We simulated exoskeleton growth of regular sea urchins as a case study for the implementation of a general computational platform allowing different approaches to data representation and modeling. Figure 3 shows different stages of the skeleton growth simulation. Starting with a circular graph with ten

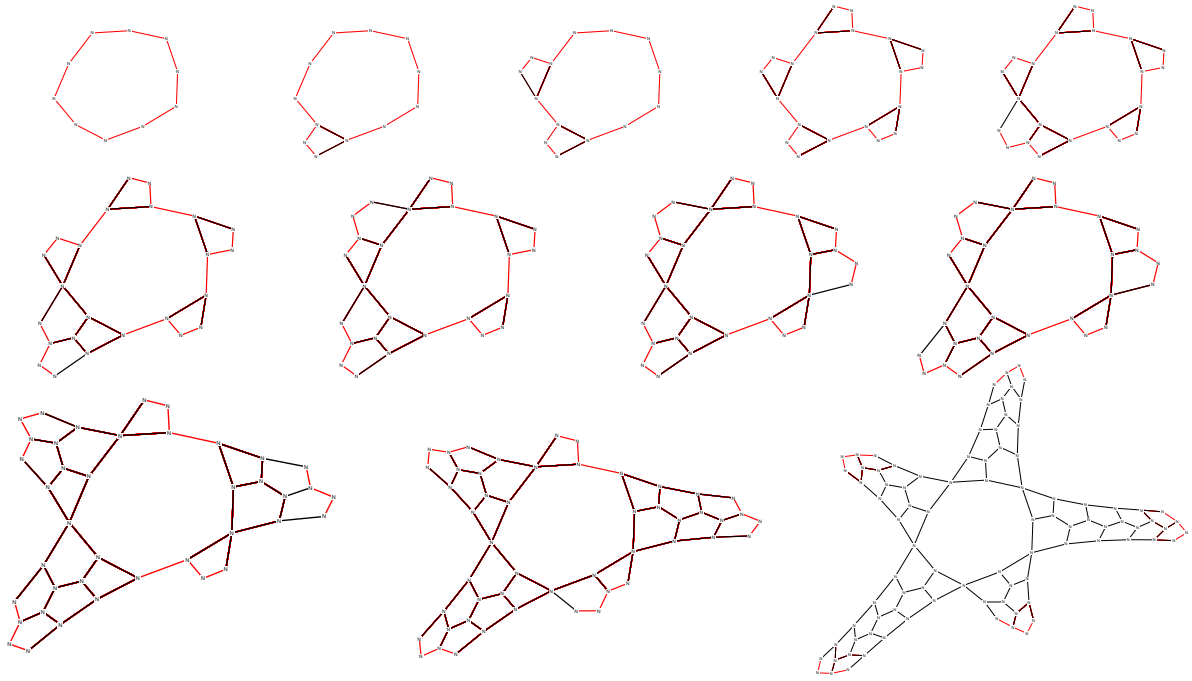

Figure 3: Using graph rewriting for the simulation of sea urchin exoskeletons (see text for details).

nodes, three rules are necessary. `Rule1` is applied only once, it attaches the first *initial plate* (a plate that starts a column) to the circular graph. (The first two graphs in Fig. 3 demonstrate the application of `Rule1`.) `Rule2` attaches an initial plate next to an already existing initial plate. (The fourth graph in Fig. 3 is the result of four applications of `Rule2`.) `Rule3` attaches pentagonal plates to the columns. As the selection of the column to which the new plate will be attached happens – as a property of the graph rewriting engine – at random, the column growth fluctuates (bottom left graph in Fig. 3). However, with an increasing number of rewriting steps this effect balances out and a mostly regular skeleton graph emerges (bottom right graph in Fig. 3). The application priority order of the rules is $2 > 1 > 3$.

5 Outlook

We hope to explore in detail the relation between combinatorial rules of change in graph structures and empirically determined frequencies of actual ontogenetic and evolutionary transitions across species. This approach allows a way of constructing new, more general theoretical morphological spaces for the comparison of the spectrum of potential forms with the actual forms realized in development and evolution. We also are interested in combining this level of simulation with models of the genome, regulatory, or transcription networks.

Fig. 2 shows how a simple simulation reproduces the developmental phenomenon heterochrony: a change of priority in the growth rules completely changes and potentially destroys the bodyplan. We are investigating how systems can be evolved that are robust against this kind of mutation.

Acknowledgements. This work is supported in part by the DFG Bioinformatics Initiative BIZ-6/1-2, the Graduiertenkolleg “Wissensrepräsentation” at the University of Leipzig, and the FWF projects P-13887-MOB and P15893-N04.

References

[1] L. Ancel and W. Fontana. Plasticity, evolvability and modularity in RNA. *J. of Exp. Zoology (Molecular and Developmental Evolution)*, 288:242–283, 2000.

[2] A. Babajide, I. L. Hofacker, M. J. Sippl, and P. F. Stadler. Neutral networks in protein space: A computational study based on knowledge-based potentials of mean force. *Folding & Design*, 2:261–269, 1997.

[3] W. Banzhaf. On the dynamics of an artificial regulatory network. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, and J. Ziegler, editors, *Advances in Artificial Life*, volume 2801 of *Lecture Notes in Computer Science*, pages 217–227, Heidelberg, Germany, 2003. Springer-Verlag. 7th European Conference, ECAL 2003, Dortmund, Germany, September 14-17, 2003, Proceedings.

[4] L. J. Croft, M. J. Lercher, M. J. Gagen, and J. S. Mattick. Is prokaryotic complexity limited by accelerated growth in regulatory overhead? *submitted to Proc. Natl. Acad. Sci. USA*, 2003. arXiv.org q-bio.MN/0311021.

[5] H. D¨orr. *Efficient Graph Rewriting and Its Implementation*. Springer-Verlag, Berlin Heidelberg, 1995.

[6] D. Drasdo and M. Loeffler. Individual-based models on growth and folding in one-layered tissues: Intestinal crypts and early development. *Nonlinear Analysis*, 47:245–256, 2001.

[7] P. Eggenberg. Evolving morphologies of simulated 3D organisms based on differential gene expression. In P. Husbands and I. Harvey, editors, *4th European Conference on Artificial Life*, pages 205–213, Cambridge, MA, 1997. The MIT Press/Bradford Books.

[8] W. Fontana and L. W. Buss. What would be conserved if 'the tape were played twice'? *Proc. Natl. Acad. Sci. USA*, 91:757–761, 1994.

[9] W. Fontana and P. Schuster. Continuity in Evolution: On the Nature of Transitions. *Science*, 280:1451–1455, 1998.

[10] N. Geard and J. Wiles. Structure and dynamics of a gene network model incorporating small RNAs. In *Proceedings of the 2003 Congress on Evolutionary Computation*, Canberra, Australia, 2003.

[11] The gml language. The `GML` language allows one to attribute arbitrary information to graphs, their nodes, and their edges. It can therefore be used to emulate almost every other data format. `http://infosun.fmi.uni-passau.de/Graphlet/GML/`.

[12] `Haskell`. is a general purpose, purely functional programming language. `Haskell` compilers are freely available for almost any computer. `http://www.haskell.org/`.

[13] M. A. Huynen, P. F. Stadler, and W. Fontana. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397–401, 1996.

[14] F. Jacob and J. Monod. On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.*, 26:193–211, 1961.

[15] S. A. Kauffman. *The Origin of Order: Self-Organization and Selection in Evolution*. Oxford University Press, Oxford, UK, 1993.

[16] A. D. Keefe and J. W. Szostak. Functional proteins from a random-sequence library. *Nature*, 410:715–718, 2001.

[17] R. C. Lewontin. *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York, 1974.

[18] J. S. Mattick and M. J. Gagen. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.*, 18(9):1661–1630, 2001.

[19] G. R. McGhee, Jr. *Theoretical Morphology*. Columbia University Press, New York, 1999.

[20] P. Nelson, M. Kiriakidou, A. Sharma, E. Maniataki, and Z. Mourelatos. The microRNA world: small is mighty. *Trends Biochem. Sci.*, 28:534–540, 2003.

[21] T. Reil. Dynamics of gene expression in an artificial genome – inplications for biological and artificial ontogeny. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *5th European Conference on Artificial Life*, volume 1674 of *Lecture Notes in Computer Science*, pages 457–466, Berlin, 1999. Springer-Verlag.

[22] E. A. Schultes and D. P. Bartel. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, 289:448–452, 2000.

[23] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, 255:279–284, 1994.

[24] S. Spiegelman. An approach to experimental analysis of precellular evolution. *Quart. Rev. Biophys.*, 4:213–253, 1971.

[25] B. M. R. Stadler, P. F. Stadler, G. Wagner, and W. Fontana. The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.*, 213:241–274, 2001.

[26] G. P. Wagner and L. Altenberg. Complex adaptations and the evolution of evolvability. *Evolution*, 50:967–976, 1996.

[27] G. Weberndorfer, I. L. Hofacker, and P. F. Stadler. On the evolution of primitive genetic codes. *Origins Life Evol. Biosph.*, 33:491–514, 2003.