Progress in Biophysics and Molecular Biology xxx (2012) 1–9

Contents lists available at SciVerse ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Original research

Driving developmental and evolutionary change: A systems biology view

Jonathan Bard*

Department of Physiology, Anatomy & Genetics, University of Oxford, UK

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Embryonic development
Evolution
Genomic module
Mutation
Network
Process

ABSTRACT

Embryonic development is underpinned by ~50 core processes that drive morphogenesis, growth, patterning and differentiation, and each is the functional output of a complex molecular network. Processes are thus the natural and parsimonious link between genotype and phenotype and the obvious focus for any discussion of biological change. Here, the implications of this approach are explored. One is that many features of developmental change can be modeled as mathematical graphs, or sets of connected triplets of the general form <*noun*><*verb*><*noun*>. In these, the verbs (edges) are the outputs of the processes that drive change and the nouns (nodes) are the time-dependent states of biological entities (from molecules to tissues). Such graphs help unpick the multi-level complexity of developmental phenomena and may help suggest new experiments. Another comes from analyzing the effect of mutation that lead to tinkering with the dynamic properties of these processes and to congenital abnormalities; if these changes are both inherited and advantageous, they become evolutionary modifications. In this context, protein networks often represents what classical evolutionary genetics sees as *genes*, and the realization that traits reflect the output processes of complex networks, particularly for growth, patterning and pigmentation, rather than anything simpler clarifies some problems that the evolutionary synthesis of the 1950s has found hard to solve. In the wider context, most processes are used many times in development and cooperate to produce tissue modules (bones, branching duct systems, muscles etc.). Their underlying generative networks can thus be thought of as genomic modules or subroutines.

1. Introduction

All biological systems are complicated, but only the brain begins to match the complexity of the early embryo. If one watches time-lapse movies of developing embryos such as those of the frog, zebrafish and *Caenorhabditis elegans*, one sees an apparently seamless and simple transition of an egg to an early embryo with its various parts. Beneath this elegant exterior, however, is a host of activities: in the mouse embryo, for example, the volume roughly quadruples and the number of named tissues doubles (from ~75 to ~150) over the 12 h between embryonic days 7 and 7.5 (Kaufman and Bard, 1999). The early embryo is a very busy place where many diverse events happen simultaneously and the only constancy is change.

When one starts to think about the principles that guide embryogenesis, one might be given the impression that the major achievement of decades of developmental biology research is merely to have shown that the development of a particular simple tissue (defined as a group of cells with the same phenotype) depends partly on its parent tissue (lineage) and partly on its neighbors (signaling). While the development of an embryo can be understood with hindsight, there are no top-level rules that predict what might happen in a particular tissue and no bottom-up rules which, for example, allow our knowledge of a tissue's gene-expression profile to predict that tissue's future. Unlike physics, there is no general theory for development and this is because it has no elementary particles and no obvious laws. Attempts have been made to show that there are the biological equivalents of thermodynamics and statistical mechanics in molecular populations (Goodwin, 1963), but they have produced little, other than showing that oscillations in their numbers are inevitable.

In the first half of the last century there was a fair amount of general conceptualizing about embryogenesis, and the dominant theoretician was certainly C.H. Waddington who invented much of the language of the subject (Waddington, 1975; Bard, 2008). Much of his work was on the nature of what we now call molecular genetics, but, because today's knowledge is far more useful than yesterday's thinking, most of this work is really only of historical interest now. Such theorizing is less common today, perhaps because we now know how very complicated development is, but we still need to have some insight into the principles of systems developmental biology, if only to understand the subject more clearly and how things go wrong when congenital abnormalities form. Furthermore, what applies to development also applies to

* Tel.: +44 (0) 1865 315185.
E-mail address: j.bard@ed.ac.uk.

0079-6107/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pbiomolbio.2012.09.006

evolution. As has been realized for many years (at least since Goldschmidt, 1927), evolutionary change in anatomical structures derives primarily from developmental change generated by germline mutations. Thus, when one starts to think about the conceptual foundations of developmental biology one is immediately required to think in a much wider context.

Much of contemporary theoretical work in the area focuses on modeling the dynamics of specific biochemical networks, work that started with the classic paper of Turing (1952) on molecular pattern formation, a theory that explains, for example, the origins of vertebrate skin patterns but for which there is as yet no experimental evidence (Bard, 1981; Barrio et al., 1999). Today, there is a considerable amount of research in this area (for review see the issue of *Current Opinion in Genetics & Development* edited by Barkai and Perrimon (2011)), mainly based on using differential equation formalisms, and there have been impressive successes in modeling a few phenomena such as signaling pathways (e.g. Witt et al., 2011), *Drosophila* segmentation (Ingolia, 2004) and somitogenesis (Goldbeter and Pourquié, 2008), but general principles that help understand development have yet to emerge. This is mainly because the activity of a signaling pathway is usually only the first step in a very complicated story, but also because the more general approaches of systems biology are not yet helpful. Noble's principles (2008), that there are no genomic programs and that causation is distributed and multi-level, certainly apply to development, but they don't really capture either the complexity or the temporal urgency of change.

The key starting point of any general thinking about development thus has to be about how change is driven, something that involves the integration of events taking place at levels that extend from gene expression, through networks and the processes that that they drive which in turn lead to changes in cell types and the production of new structures. This paper starts by exploring these processes and the molecular networks that generate them. I then show how it is possible to represent many facets of the multi-level complexity that underpins structural change within what is known as a mathematical graph (details are given in the Appendix). The advantages of this representation are then considered, together with some difficulties and limitations. The last major section discusses how this focus on processes allows us to start to fuse the two major strands of evolutionary thinking: traditional approaches centering around phenotypes (the evolutionary synthesis; Mayr and Provine, 1980) and modern work analyzing genomic changes (e.g. Shapiro, 2011). The paper ends with a brief consideration of the concepts underpinning systems developmental biology.

2. Processes

Development involves the integration of events at all levels from genome sequences through to tissues. Table 1 gives some idea of numbers of components at the various levels for the mouse and human; Fig. 1 indicates how these events are integrated. The figures for protein-coding genes and proteins are well known; the number of developmental networks and output actions comes partly from Gilbert (2010) and partly from analyzing in detail the changes that take place as humans develop over the first seven weeks (www.obofoundry.org/cgi-bin/detail.cgi?id=human-dev-anat-abstract2). It turns out that the number of simple tissues that have formed in the human embryo by about 7 weeks when the embryo is ~2 cm in length is more than two thousand, a figure that excludes the minor blood vessels and secondary nerves.

The evidence to date on the development of any organisms suggests that it is driven by surprisingly few processes, each of which is underpinned by a protein network (Tables 1 and 2). There are ~10 signal-activated networks and these in turn activate 10–20 spatial patterning networks (details are unclear here), 5–10 pathways associated with proliferation, ~3 apoptosis networks, 5–10 morphogenetic networks, and a hierarchy of differentiation pathways. The number of high-level differentiation pathways is less obvious because major cell types have subtypes, but one pointer here comes from the options available to neural crest cells. These include, mesenchymal cells (bone, muscle, cartilage, fibroblasts) epithelia of various sorts, neurons and neuron-support cells and melanocytes, but not the other major lineage of blood cells and their many subtypes. There are thus perhaps ~10 main cell differentiation routes. Taking a broad brush to the topic, there are ~50 major processes that underpin development (perhaps ~60 if we allow for the possibility that a few more will be discovered), with some having several outputs (Fig. 2). It is a major achievement of the last 20 or so years to have shown how closely related are equivalent networks across the phyla (Gilbert, 2010).

Table 1
Levels and numbers in the late mammalian embryo.

Protein-coding genes	~20 K
Proteins	~70 K
Developmental Networks generating output processes	~60
Simple tissues	~10 K

Here, it should be emphasized that, although the development of each complex, multicellular animal requires many of these processes, evolutionary change means that the fine details of the networks and their outputs vary from organism to organism and their analysis is of course the bread-and-butter work of developmental biologists. It should also be pointed out that the repertoire of developmental networks (Table 2) excludes the many more networks that "run" the biochemical, physiological and neurological systems. Nevertheless, if there is any underlying simplicity to be found in developmental biology, it centres around a basic set of molecular networks[1] whose outputs are the processes that drive embryogenesis (Fig. 1).

Where they are known, the number of proteins in a network ranges from ~10 (e.g. the hedgehog signaling pathway) to >50 (e.g. the Rho-GTPase network that regulates much of cell morphology and movement and the EGF network that frequently activates proliferation. For further details, see http://www.sabiosciences.com/pathwaycentral.php and Fig. 2). Elucidating the components and the organization of these networks has been a triumph of the last decade of research in molecular genetics. It has however proved far more difficult to understand how they work or, apart from the relatively simple signal transduction pathways, to model their behavior qualitatively, let alone quantitatively. Some progress has been made in identifying functional motifs where small groups of proteins co-operate to produce a functional unit such as a the feed-forward motif (Alon, 2007) or how several proteins comprise a single structural unit by attaching to further proteins that function as docking stations (Simister and Feller, 2012). In general, however, it is still hard to see how they function dynamically, except in those rare cases where it has been shown that the topology of the network imposes a robust output (Ingolia, 2004).

The situation is made more complicated by the fact that individual variation in DNA sequences may lead to minor alterations in protein structure and hence to changes in functional rate constants. The evolutionary implications of this are discussed below; here, it is merely pointed out that the networks have to be robust enough to allow such minor variation to occur while still producing a reliable

[1] What are called networks here are more commonly called pathways (many are available at http://www.sabiosciences.com/pathwaycentral.php?a). The former is the preferable term because these assemblages of proteins often include alternate routes and endpoints.

Please cite this article in press as: Bard, J., Driving developmental and evolutionary change: A systems biology view, Progress in Biophysics and Molecular Biology (2012), http://dx.doi.org/10.1016/j.pbiomolbio.2012.09.006