Spring 2019 – Epigenetics and Systems Biology
Discussion Outline (Systems Biology)
Michael K. Skinner – Biol 476/576
Weeks 1 and 2 (January 17)

Systems Biology

Primary Papers

1. Kitano H. (2002) Nature 240:206-210
2. Morelli, et al. (2012) Science 336:187-191
3. Sarma, et al. (2018) Philosophical Transactions B 373:20170382

Discussion

Student 1 - Ref #1 above
-What are simulation and in silico experiments?
-What are scale free networks?
-How can this computational approach help medicine?

Student 2 - Ref #2 above
-What are patterning strategies?
-What is mechanical deformation?
-How are gene networks involved?

Student 3 - Ref #3 above
-What is open worm project and big science?
-How is computational simulation involved?
-What are the insights provided and computational approaches?

Computational systems biology

Hiroaki Kitano

Sony Computer Science Laboratories Inc., 3-14-13 Higashi-gotanda, Shinagwa, Tokyo 141-0022, ERATO Kitano Symbiotic Systems Project, Japan Science and Technology Corporation, and The Systems Biology Institute, Suite 6A, M31, 6-31-15 Jingu-mae, Shibuya, Tokyo 150-0001, School of Fundamental Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan, and Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125, USA (e-mail: kitano@csl.sony.co.jp)

To understand complex biological systems requires the integration of experimental and computational research — in other words a systems biology approach. Computational biology, through pragmatic modelling and theoretical exploration, provides a powerful foundation from which to address critical scientific questions head-on. The reviews in this Insight cover many different aspects of this energetic field, although all, in one way or another, illuminate the functioning of modular circuits, including their robustness, design and manipulation. Computational systems biology addresses questions fundamental to our understanding of life, yet progress here will lead to practical innovations in medicine, drug discovery and engineering.

It is often said that biological systems, such as cells, are 'complex systems'. A popular notion of complex systems is of very large numbers of simple and identical elements interacting to produce 'complex' behaviours. The reality of biological systems is somewhat different. Here large numbers of functionally diverse, and frequently multifunctional, sets of elements interact selectively and nonlinearly to produce coherent rather than complex behaviours.

Unlike complex systems of simple elements, in which functions emerge from the properties of the networks they form rather than from any specific element, functions in biological systems rely on a combination of the network and the specific elements involved. For example, p53 (a 393-amino-acid protein sometimes called 'the guardian of genome') acts as tumour suppressor because of its position within a network of transcription factors. However, p53 is activated, inhibited and degraded by modifications such as phosphorylation, dephosphorylation and proteolytic degradation, while its targets are selected by the different modification patterns that exist; these are properties that reflect the complexity of the element itself. Neither p53 nor the network functions as a tumour suppressor in isolation. In this way, biological systems might be better characterized as symbiotic systems.

Molecular biology has uncovered a multitude of biological facts, such as genome sequences and protein properties, but this alone is not sufficient for interpreting biological systems. Cells, tissues, organs, organisms and ecological webs are systems of components whose specific interactions have been defined by evolution; thus a system-level understanding should be the prime goal of biology. Although advances in accurate, quantitative experimental approaches will doubtless continue, insights into the functioning of biological systems will not result from purely intuitive assaults. This is because of the intrinsic complexity of biological systems. A combination of experimental and computational approaches is expected to resolve this problem.

A two-pronged attack

Computational biology has two distinct branches: knowledge discovery, or data-mining, which extracts the hidden patterns from huge quantities of experimental data, forming hypotheses as a result; and simulation-based analysis, which tests hypotheses with *in silico* experiments, providing predictions to be tested by *in vitro* and *in vivo* studies.

Knowledge discovery is used extensively within bioinformatics for such tasks as the prediction of exon–intron and protein structure from sequence[1], and the inference of gene regulatory networks from expression profile[2–4]. These methods typically use predictions based on heuristics, on statistical discriminators that often involve sophisticated approaches (such as hidden Markov models) and on other linguistic-based algorithms (see review in this issue by Searls, pages 211–217).

In contrast, simulation attempts to predict the dynamics of systems so that the validity of the underlying assumptions can be tested. Detailed behaviours of computer-executable models are first compared with experimental observation. Inconsistency at this stage means that the assumptions that represent our knowledge on the system under consideration are at best incomplete. Models that survive initial validation can then be used to make predictions to be tested by experiments, as well as to explore questions that are not amenable to experimental inquiry.

Although traditional bioinformatics has been used widely for genome analysis, simulation-based approaches have received little mainstream attention. This is now changing. Current experimental molecular biology is now producing the high-throughput quantitative data needed to support simulation-based research. Combined with rapid progress of genome and proteome projects, this is convincing increasing numbers of researchers of the importance of a system-level approach[5]. At the same time, substantial advances in software and computational power have enabled the creation and analysis of reasonably realistic yet intricate biological models.

There are still issues to be resolved, but computational modelling and analysis are now able to provide useful biological insights and predictions for well understood targets such as bifurcation analysis of the cell cycle[6,7], metabolic analysis[8,9] or comparative studies of robustness of biological oscillation circuits[10].

It is crucial that individual research groups are able to exchange their models and create commonly accepted repositories and software environments that are available to all. Systems Biology Markup Language (SBML; http://www.sbml.org/), CellML (http://www.cellml.org/) and the Systems Biology Workbench are examples of efforts that aim to form a *de facto* standard and open software platform for modelling and analysis[11,12]. These significantly increase the value of the new generation of databases concerned with biological pathways, such as the Kyoto

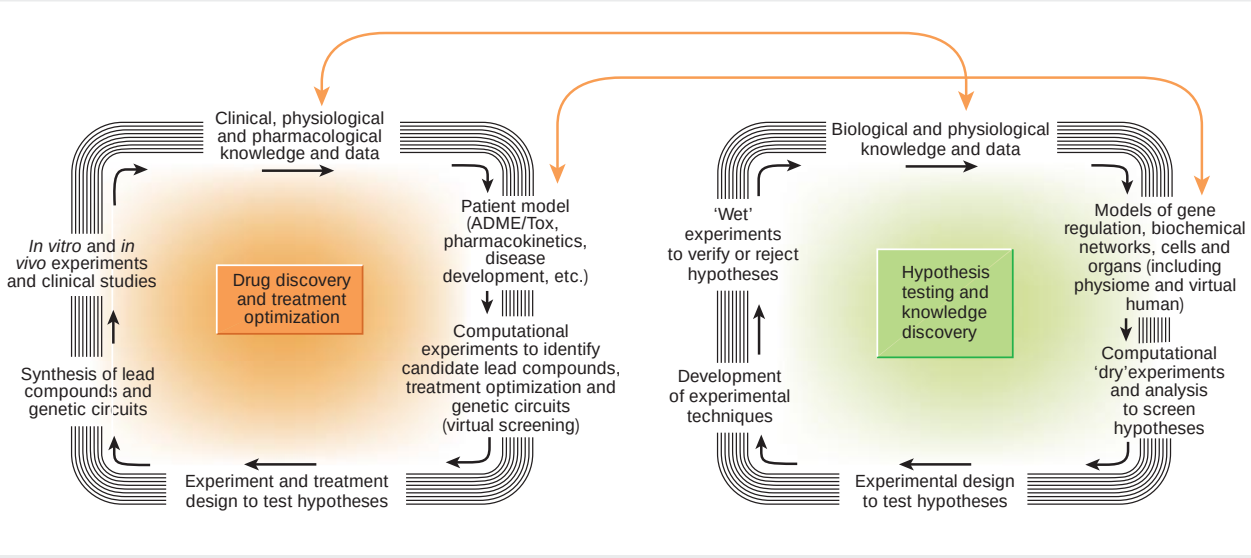

Figure 1 Linkage of a basic systems-biology research cycle with drug discovery and treatment cycles. Systems biology is an integrated process of computational modelling, system analysis, technology development for experiments, and quantitative experiments[18]. With sufficient progress in basic systems biology, this cycle can be applied to drug discovery and the development of new treatments. In the future, *in silico* experiments and screening of lead candidates and multiple drug systems, as well as introduced genetic circuits, will have a key role in the 'upstream' processes of the pharmaceutical industry, significantly reducing costs and increasing the success of product and service development.

Encyclopedia of Genes and Genomes (KEGG)[13], Alliance for Cellular Signaling (AfCS)[14] and Signal Transduction Knowledge Environment (STKE)[15], by enabling them to develop machine-executable models, rather than mere human-readable forms.

Such changes are fuelling a renewed interest in a system-level approach to biology, but we should not forget that this is an area with a long history[16,17], rooted as much as anywhere in classical physiology (see review in this issue by Buchman, pages 246–251). However, the close linkage between system-level understanding and molecular-level knowledge was made possible only by the recent progress in genomics and proteomics. The approach attempts to understand biological systems as systems, specifically targeting the identification of their structures and dynamics, and the establishment of methods to control cellular behaviours by external stimuli and to design genetic circuits with desired properties. These aims will be achieved only by combining computation, system analysis, new technologies for comprehensive and quantitative measurements, and high-throughput quantitative experimental data[18,19].

Multiple faces of robustness

Among various scientific questions, one issue receiving considerable attention is how robustness is achieved and how it evolves within various aspects of biological systems. Robust systems maintain their state and functions against external and internal perturbations, and robustness is an essential feature of biological systems, having been studied since the earliest attempts at a system-oriented view (for example, Cannon's homeostasis and Weiner's cybernetics[16]). Biological systems have been found to be robust at a variety of levels from genetic switches to physiological reactions (see review in this issue by Buchman, pages 246–251).

Robust systems are both relatively insensitive to alterations of their internal parameters and able to adapt to changes in their environment. In highly robust systems, even damage to their very structure produces only minor alterations in their behaviour. Such properties are achieved through feedback, modularity, redundancy and structural stability.

A variety of feedback and feed-forward control is observed throughout biology. For example, integral feedback is central to bacteria chemotaxis[20–22]. And p53-based cell-cycle arrest displays what is known in the engineering field as 'bang-bang control', a subtype of feedback control. Damage to DNA is sensed by proteins such as ATM (for ataxia telangiectasia mutated, named after a disease in which this enzyme is mutated) and DNA-dependent protein kinase, which activate the p53 protein. Active p53 then transactivates p21, which results in G1 arrest; this state is released when DNA damage is repaired, thus forming a feedback loop.

Cells themselves provide the most obvious form of biological modularity by physically partitioning off biochemical reactions. However, biochemical networks within cells also form modular compartments isolated by spatial localization[23], anchoring of proteins to plasma membranes and by dynamics.

Cells also provide redundancy, with many autonomous units carrying out identical roles. But redundancy also appears at other levels by having multiple genes that encode similar proteins, or multiple networks with complementary functions. For example, *Per1*, *Per2* and *Per3* genes encode proteins in the circadian oscillator, but knock-out of one or two of these produces no visible phenotype. The *Cln* gene family form redundant pairs for the cell cycle[24]. The stringent response of *Escherichia coli* activates alternative metabolic dynamics depending upon the availability of lactose and glucose[25].

Structurally stable network configurations increase insensitivity to parameter changes, noise and minor mutations. For example, elegant experiments on the archetypal genetic switch — the lambda phage decision circuit — have shown it to be robust against changes in binding affinity of promoters and repressors; its stable switching action arises from the structure of its network, rather than the specific affinities of its binding site[26]. Additionally, a number of networks for biological oscillations and transcriptional regulations have been shown to be tolerant against noise (ref. 27; and see review in this issue by Rao and colleagues, pages 231–237). But only computer simulation could have shown the degree to which the gene regulatory networks for segmentation during *Drosophila* embryogenesis remain robust over a large range of kinetic parameters[28,29].

The robustness of a system is not always to an organism's advantage. Cancer cells are extremely robust for their own growth and survival against various perturbations. They continue to proliferate, driven by the engine of the cell cycle, eliminating

communication with their external environment, thus making it insensitive against external perturbations. In addition, many anti-cancer drugs are rendered ineffective by the normal functioning of a patient's body, including defence systems such as the metabolism of xenobiotics (most notably by cytochrome P450), the brain–blood barrier, and the dynamics of gene regulatory circuits, which can adjust the concentration of drug targets through feedback mechanisms and redundancy. To establish treatments that move patients from a stable but diseased state to a healthy one will require an in-depth, system-level understanding of biological robustness.

Although the general principles of robust systems are well established, there remain a number of unresolved issues concerning their evolution and execution in specific biological systems, and how they can be manipulated or designed. Control theory has been used to provide a theoretical underpinning of some robust systems, such as adaptation through negative feedback[21]. However, this approach has limitations. For example, current control theory assumes that target values or statuses are provided initially for the systems designer, whereas in biology such targets are created and revised continuously by the system itself. Such self-determined evolution is beyond the scope of current control theory.

No free lunch

Although robustness is critical in assuring the survival of a biological system, it does not come without cost. Carlson and Doyle emphasize the "robust, yet fragile" nature of complex systems exhibiting highly optimized tolerance[30,31]. Systems designed or evolved to be robust against common or known perturbations can often be fragile to new perturbations.

Another view on the vulnerability of complex network comes from a statistical perspective[32–34]. Comparative studies on robustness of large-scale networks show that scale-free networks (also known as 'small world' or Erdös–Rényi networks) are more robust than randomly connected networks against random failure of their components[34]. However, scale-free networks are more vulnerable against malfunction of the few highly connected nodes that function as hubs.

Scale-free networks can form by growth such that new nodes are connected preferentially to nodes that are already highly connected. Barabasi and colleagues claim that protein–protein interaction networks, which constitute the protein universe (see review in this issue by Koonin and colleagues, pages 218–223), are scale-free[32,35] and that mutations in highly connected proteins are more likely to be lethal than are mutations in less-connected nodes[33]. Although they estimated connectivity from yeast two-hybrid data, which are notoriously noisy, this hypothesis is intuitively attractive. For example, the p53 protein is one of the most connected hubs in the protein universe, and its mutations cause serious damage to cellular functions, particularly in repair of DNA damage and tumour suppression[36].

Nevertheless, some of the claims for scale-free networks are still controversial[37], and evidence for mechanisms leading to preferential attachment in biological systems remains equivocal. Furthermore, yeast two-hybrid assays produce many false-positive outcomes, and the current hand-crafted pathway maps may be heavily biased towards connection to functionally important genes simply because these have been popular targets for research.

Even when these shortcomings are surpassed, such statistics-based theories — despite providing insights on macroscopic properties of the network — will still have difficulty making predictions about specific interactions. It is analogous to telling a stock-market investor that "one in 50 companies will go bankrupt", advice that is of little help if you are unable to identify which one. The challenge for statistical theories is to identify how they can be linked to specific behaviours and so make useful predictions.

Design patterns of functional modules

Just as the principles behind robust networks can be classified into several types, so too can the various functional circuits or modules from which they are assembled, such as genetic switches, flip-flops, logic gates, amplifiers and oscillators. Good examples come from the mechanisms of biochemical oscillations (see review in this issue by Goldbeter, pages 238–245), which have been the focus of numerous groups[38–41]. These studies have facilitated their classification into several schemes, such as substrate-depletion oscillators, positive feedback loops, the Goodwin oscillator and time-delayed negative feedback oscillators[41]. Similar attempts have also been made for other functional networks. Jordan and colleagues have identified various examples of multitasking in signal transduction[42]; Bhalla and Iyengar reported several circuits that may function as temporal information stores (that is, memory devices)[43]; and Rao and colleagues have uncovered several circuits that mitigate the effect of noise and exploit it for specific functions (see review in this issue, pages 231–237).

Although these functional networks have analogues in electronic and process engineering, they have been formed by evolution, which makes it unlikely that any kind of 'first principle' underlies their design. However, a set of principles can be envisaged and identified through studying the structure and function of biological circuits, and their origin at the system level[44–46]. What are their basic functional building blocks? What are their dynamical properties and operating principles? How has each module evolved? And how can they be adapted or designed for alternative applications?

Recently, a systematic, high-throughput computational study was carried out by Shen-Orr and colleagues, which identified common motifs in the gene regulatory networks of *E. coli* using the RegulonDB database[47]. They found that feed-forward loops, single-input modules and dense overlapping regulons appeared frequently. While this study only used a gene regulation database, this type of approach can be augmented to include protein–protein and protein–DNA interactions to systematically identify network design patterns from large-scale data.

Such data, combined with function-driven identification of circuit patterns, will allow the creation of a large repository of functional biological networks, so enabling the systematic analysis of design patterns and their evolution. We already know of cases where the same circuit patterns and homologous genes produce similar system behaviours, but with unrelated physiological outcomes. We also know of cases where the same circuit patterns use different sets of genes to attain similar system behaviours, and where identical functions are achieved with degenerate paths involving different circuit patterns and different genes[46]. More systematic surveys will be needed to determine how many evolutionary conserved circuits exist, in what functions and how they relate to the evolution of genes. It may be that functional circuits should be considered the units of evolution.

Systems drug and treatment discovery

The systems biology approach, with its combination of computational, experimental and observational enquiry, is highly relevant to drug discovery and the optimization of medical treatment regimes for individual patients. Although the analysis of individual single nucleotide polymorphisms is expected to reveal individual genetic susceptibilities to all forms of pathological condition, it may be impossible to identify such relationships when complex interactions are involved.

Consider a hypothetical example where variations of gene A induce a certain disease. Susceptibility relationships may not be apparent if circuits exist to compensate for the effects of the variability. Polymorphisms in gene A will be linked to disease susceptibility only if these compensatory circuits break down for some reason. A more mechanistic, systems-based analysis will be necessary to elucidate more complex relationships involving multiple genes that may create new opportunities for drug discovery and treatment optimization.

Computer simulation and analysis, along with traditional bioinformatics approaches, have frequently been proposed to significantly increase the efficiency of drug discovery[48–50]. At present, empirical ADME/Tox (absorption distribution metabolism excretion/toxicity) and pharmacokinetic predictions have been used with some success.

For example, a human intestinal absorption model based on correlations between the passive permeation measurement of over 300 compounds and known structural features, such as hydrogen-bond donors, hydrogen-bond acceptors and molecular weight, has been used to predict the absorption of novel compounds by the human intestine[51]. However, such models are not easily converted for use in other situations and they often require extensive data sets in order to address specific questions. What is needed are reliable, mechanism-based ADME/Tox and pharmacokinetic models[52–56], built on molecular-level models of cells, that are more easily transferable and accountable than are traditional, empirical, quantitative structure–activity relations.

Scaling up

So far, most systems biology simulations have tended to target relatively small sub-networks within cells, such as the feedback circuit for bacteria chemotaxis[20,21], the circadian rhythm[57,58], parts of signal-transduction pathways[43,59], simplified models of the cell cycle[7,60,61] and red blood cells[62–64]. Notable larger simulations have attempted to model bacterial metabolic networks for analysis of metabolic control[62,63] and flux balance[8,65], but these deal with steady-state rather than dynamic behaviour. Recently, research has begun on larger-scale simulations. At the level of the biochemical network, simulation of the epidermal growth factor (EGF) signal-transduction cascade has been carried out. The simulation involves over 100 equations and kinetic parameters and will be used to predict complex behaviours of the pathway, as well as to identify roles of external and internal EGF receptors[59]. The physiome project is an ambitious attempt to create virtual organs that represent essential features of organs *in silico*[66,67]. Simulation of the heart was one of the early attempts in this direction, integrating multiple scales of models from genetics to physiology[68]. Even whole-patient models for specific disease, such as obesity and diabetes, are being developed for prediction of disease development and drug discovery.

Building a full-scale patient model, or even a whole-cell or organ model, is a challenging enterprise. Multiple aspects of biological processes have to be integrated and the model predictions must be verified by biological and clinical data, which are at best sparse for this purpose. Integrating heterogeneous simulation models is a non-trivial research topic by itself, requiring integration of data of multiple scales, resolutions and modalities.

Simulation often requires integration of multiple hierarchies of models that are orders of magnitude different in terms of scale and qualitative properties (for example, gene regulations, biochemical networks, intercellular communications, tissue, organ and patient). Although some processes can be modelled by either stochastic computation or differential equations alone, many require a combination of both methods. But some biochemical processes take place within a millisecond whereas others can take hours or days. Additionally, biological processes often involve the interaction of different types of process, such as biochemical networks coupled to protein transport, chromosome dynamics, cell migration or morphological changes in tissues. Although biochemical networks may be reasonably modelled using differential equations and stochastic simulation, many cell biological phenomena require calculation of structural dynamics, deformation of elastic bodies, spring-mass models and other physical processes.

Nevertheless, development of precision models and their applications to ADME/Tox models are expected to revolutionize the process of drug discovery by providing a capability for multiple-target identification and high-throughput virtual screening of compounds. Furthermore, target identification using cellular models may provide desirable structures for candidate compounds by applying multiple constraints to parallel virtual screening[54], rationalizing drug discovery into a more systematic process (Fig. 1).

Systems therapy

Surpassing its scope for efficient improvements in the current paradigm of drug discovery and treatment, the introduction of a system-oriented view may drastically change the way treatments are conducted. Two somewhat speculative scenarios illustrate these opportunities.

Consider a feedback compensation circuit involving a drug target protein. Changes in the concentration of the protein resulting from drug administration may be neutralized by feedback control. High dosages of drugs will need to be administered to overcome this compensation mechanism, but this could produce serious side effects. Alternatively, small dosages of drugs could mitigate the feedback mechanism, so that the effect on the target protein will not be neutralized. Considering the p53 system, if there is abnormal overexpression of MDM2 (a protein that regulates p53), simply increasing p53 transcription may not restore the system to normal, as the excessive MDM2 protein will quickly ubiquitinate p53, targeting it for destruction. Additionally, p53 itself transactivates MDM2. MDM2 activity must be suspended or reduced to a normal level, at least temporarily, to make p53 stimulation effective in inducing cell-cycle arrest or apoptosis. The highly effective administration of multiple drug regimes can be accomplished only with a system-level analysis of the dynamics of gene regulatory circuits.

A far more futuristic approach proposes the introduction of functional genetic circuits to control cellular dynamics *in vivo* (see review in this issue by Hasty and colleagues, pages 224–230). Already, a set of basic functional circuits, such as oscillators and toggle switches, has been constructed and its viability confirmed in *E. coli* (refs 69–71; and see review by Hasty and colleagues). Computer simulation and comprehensive analysis will be needed to ensure that such circuits function as intended and do not result in significant side-effects. In the future, perhaps a genetic circuit can be devised to sense the level of p53 protein when DNA is damaged and switch on circuits to further increase transcription of p53.

The application of systems biology to medical practice is the future of medicine. Its realization will see drug discovery and the design of multiple drug therapies and therapeutic gene circuits being pursued just as occurs now with modern, complex engineering products — through iterative cycles of hypothesis and simulation-driven processes (Fig. 1). Although the road ahead is long and winding, it leads to a future where biology and medicine are transformed into precision engineering. □

doi:10.1038/nature01254

1. Baldi, P. & Brunak, S. *Bioinformatics: The Machine Learning Approach* 2nd edn (MIT Press, Cambridge, MA, 2001).
2. Onami, S., Kyoda, K., Morohashi, M. & Kitano, H. in *Foundations of Systems Biology* (ed. Kitano, H.) 59–75 (MIT Press, Cambridge, MA, 2001).
3. Ideker, T. E., Thorsson, V. & Karp, R. M. in *Pac. Symp. Biocomput.* (eds Altman, R. B., Dunker, A. K., Hunter, L., Lauderdale, K. & Klein, T. E.) 305–316 (World Scientific, Singapore, 2000).
4. Ideker, T. *et al.* Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**(Suppl. 1), S233–S240 (2002).
5. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292,** 929–934 (2001).
6. Borisuk, M. T. & Tyson, J. J. Bifurcation analysis of a model of mitotic control in frog eggs. *J. Theor. Biol.* **195,** 69–85 (1998).
7. Chen, K. C. *et al.* Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* **11,** 369–391 (2000).
8. Edwards, J. S., Ibarra, R. U. & Palsson, B. O. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.* **19,** 125–130 (2001).
9. Fell, D. *Understanding the Control of Metabolism* (Portland, London, 1997).
10. Morohashi, M. *et al.* Robustness as a measure of plausibility in models of biochemical networks. *J. Theor. Biol.* **216,** 19–30 (2002).
11. Kitano, H. Standards for modeling. *Nature Biotechnol.* **20,** 337 (2002).
12. Hucka, M. *et al.* in in *Pac. Symp. Biocomput.* (eds Altman, R. B., Dunker, A. K., Hunter, L. & Klein, T. E.) 450–461 (World Scientific, Singapore, 2002).
13. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
14. Alliance for Cellular Signaling <http://www.AfCS.org/> (2002).
15. Signal Transduction Knowledge Environment <http://www.stke.org/> (2002).
16. Wiener, N. *Cybernetics: Or Control and Communication in the Animal and the Machine* (MIT Press, Cambridge, MA, 1948).
17. Bertalanffy, L. v. *General System Theory* (Braziller, New York, 1968).
18. Kitano, H. Systems biology: a brief overview. *Science* **295,** 1662–1664 (2002).
19. Kitano, H. in *Foundations of Systems Biology* (ed. Kitano, H.) 1–36 (MIT Press, Cambridge, MA, 2001).
20. Alon, U. *et al.* Robustness in bacterial chemotaxis. *Nature* **397,** 168–171 (1999).

21. Yi, T. M. *et al.* Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl Acad. Sci. USA* **97,** 4649–4653 (2000).
22. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387,** 913–917 (1997).
23. Weng, G., Bhalla, U. S. & Iyengar, R. Complexity in biological signaling systems. *Science* **284,** 92–96 (1999).
24. Levine, K., Tinkelenberg, A. & Cross, F. in *Progress in Cell Cycle Research* (eds Meijer, L., Guidet, S. & Lim Tung, H. Y.) 101–114 (Plenum, New York, 1995).
25. Chang, D. E., Smalley, D. J. & Conway, T. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol. Microbiol.* **45,** 289–306 (2002).
26. Little, J. W., Shepley, D. P. & Wert, D. W. Robustness of a gene regulatory circuit. *EMBO J.* **18,** 4299–4307 (1999).
27. Gonze, D., Halloy, J. & Goldbeter, A. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl Acad. Sci. USA* **99,** 673–678 (2002).
28. von Dassow, G. *et al.* The segment polarity network is a robust developmental module. *Nature* **406,** 188–192 (2000).
29. Eldar, A. *et al.* Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. *Nature* **419,** 304–308 (2002).
30. Carlson, J. M. & Doyle, J. Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E* **60,** 1412–1427 (1999).
31. Carlson, J. M. & Doyle, J. Complexity and robustness. *Proc. Natl Acad. Sci. USA* **99,** 2538–2545 (2002).
32. Jeong, H. *et al.* The large-scale organization of metabolic networks. *Nature* **407,** 651–654 (2000).
33. Jeong, H. *et al.* Lethality and centrality in protein networks. *Nature* **411,** 41–42 (2001).
34. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **406,** 378–382 (2000).
35. Podani, J. *et al.* Comparable system-level organization of Archaea and Eukaryotes. *Nature Genet.* **29,** 54–56 (2001).
36. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408,** 307–310 (2000).
37. Adamic, L. A., Lukose, R. M., Puniyani, A. R. & Huberman, B. A. Search in power-law networks. *Phys. Rev. E* **64,** 046135-1–046135-8 (2001).
38. Higgins, J. The theory of oscillating reactions. *Ind. Eng. Chem.* **59,** 18–62 (1967).
39. Berridge, M. J. & Rapp, P. E. A comparative survey of the function, mechanism and control of cellular oscillators. *J. Exp. Biol.* **81,** 217–279 (1979).
40. Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms* (Cambirdge Univ. Press, Cambridge, 1996).
41. Tyson, J. J. in *Computatoinal Cell Biology* (eds Fall, C. P., Marland, E. S., Wagner, J. M. & Tyson, J. J.) 230–260 (Springer, New York, 2002).
42. Jordan, J. D., Landau, E. M. & Iyengar, R. Signaling networks: the origins of cellular multitasking. *Cell* **103,** 193–200 (2000).
43. Bhalla, U. S. & Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **283,** 381–387 (1999).
44. Hartwell, L. H. *et al.* From molecular to modular cell biology. *Nature* **402,** C47–C52 (1999).
45. Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science* **295,** 1664–1669 (2002).
46. Edelman, G. M. & Gally, J. A. Degeneracy and complexity in biological systems. *Proc. Natl Acad. Sci. USA* **98,** 13763–13768 (2001).
47. Shen-Orr, S. S. *et al.* Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31,** 64–68 (2002).
48. Cascante, M. *et al.* Metabolic control analysis in drug discovery and disease. *Nature Biotechnol.* **20,** 243–249 (2002).
49. Bailey, J. E. Lessons from metabolic engineering for functional genomics and drug discovery. *Nature Biotechnol.* **17,** 616–618 (1999).
50. Bailey, J. E. Reflections on the scope and the future of metabolic engineering and its connections to functional genomics and drug discovery. *Metab. Eng,* **3,** 111–114 (2001).
51. Lipinski, C. A. *et al.* Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46,** 3–26 (2001).
52. Butina, D., Segall, M. D. & Frankcombe, K. Predicting ADME properties in silico: methods and models. *Drug Discov. Today* **7,** S83–S88 (2002).
53. Ekins, S. & Rose, J. In silico ADME/Tox: the state of the art. *J. Mol. Graph. Model.* **20,** 305–309 (2002).
54. Selick, H. E., Beresford, A. P. & Tarbit, M. H. The emerging importance of predictive ADME simulation in drug discovery. *Drug Discov. Today* **7,** 109–116 (2002).
55. Li, A. P. & Segall, M. Early ADME/Tox studies and in silico screening. *Drug Discov. Today* **7,** 25–27 (2002).
56. Ekins, S. *et al.* Progress in predicting human ADME parameters in silico. *J. Pharmacol. Toxicol. Methods* **44,** 251–272 (2000).
57. Ueda, H. R., Hagiwara, M. & Kitano, H. Robust oscillations within the interlocked feedback model of *Drosophila* circadian rhythm. *J. Theor. Biol.* **210,** 401–406 (2001).
58. Leloup, J. C., Gonze, D. & Goldbeter, A. Limit cycle models for circadian rhythms based on transcriptional regulation in *Drosophila* and *Neurospora*. *J. Biol. Rhythms* **14,** 433–448 (1999).
59. Schoeberl, B. *et al.* Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnol.* **20,** 370–375 (2002).
60. Tyson, J. J. & Novak, B. Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions. *J. Theor. Biol.* **210,** 249–263 (2001).
61. Novak, B. *et al.* Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophys. Chem.* **72,** 185–200 (1998).
62. Ni, T. C. & Savageau, M. A. Model assessment and refinement using strategies from biochemical systems theory: application to metabolism in human red blood cells. *J. Theor. Biol.* **179,** 329–368 (1996).
63. Ni, T. C. & Savageau, M. A. Application of biochemical systems theory to metabolism in human red blood cells. Signal propagation and accuracy of representation. *J. Biol. Chem.* **271,** 7927–7941 (1996).
64. Jamshidi, N. *et al.* Dynamic simulation of the human red blood cell metabolic network. *Bioinformatics* **17,** 286–287 (2001).
65. Edwards, J. S. & Palsson, B. O. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* **16,** 927–939 (2000).
66. Bassingthwaighte, J. B. Strategies for the physiome project. *Ann. Biomed. Eng.* **28,** 1043–1058 (2000).
67. Rudy, Y. From genome to physiome: integrative models of cardiac excitation. *Ann. Biomed. Eng.* **28,** 945–950 (2000).
68. Noble, D. Modeling the heart—from genes to cells to the whole organ. *Science* **295,** 1678–1682 (2002).
69. Guet, C. C. *et al.* Combinatorial synthesis of genetic networks. *Science* **296,** 1466–1470 (2002).
70. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403,** 339–342 (2000).
71. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).

Acknowledgements

I thank S. Imai, J. Doyle, J. Tyson, T.-M. Yi, N. Hiroi and M. Morohashi for their useful comments on the manuscript. This research is, in part, supported by: the Rice Genome and Simulation Project (Ministry of Agriculture), International Standard Development area of International Joint Research Grant (New Energy and Industrial Technology Development Organization (NEDO)/Japanese Ministry of Economy, Trade and Industry (METI)), Exploratory Research for Advanced Technology (ERATO) and Institute for Bioinformatics Research and Development (BIRD) program (Japan Science and Technology Corporation), and through the special coordination funds for promoting science and technology from the Japanese government's Ministry of Education, Culture, Sports, Science, and Technology.

the mRNA X and mRNA Y corresponding to protein X and protein Y, respectively. Although protein X and protein Y are coordinated for all four motifs in Fig. 3, this is not the case for their mRNA levels. This can be explained by the disparate time scales of mRNA and protein. Fast-degrading mRNA may exhibit fluctuations with a broad frequency bandwidth. Conversely, slow degradation of proteins filters out fast fluctuations but keeps slow fluctuations. Constitutively expressed mRNA X has both fast and slow fluctuations, but protein X only transmits the slow fluctuations downstream. The result is that the dynamics of mRNA X and mRNA Y are dominated by uncorrelated fast fluctuations, which overshadow their correlated slow fluctuations. On the other hand, protein X and protein Y only contain the better-correlated slow fluctuations. That is, two mRNA species can be mostly uncorrelated with one another, yet produce protein in a coordinated fashion. Gandhi *et al.* (*18*) observed such a circumstance in budding yeast, when they found very little correlation between pairs of transcripts that encode coordinated proteins of the same protein complex, including proteasome and RNA polymerase II subunits. They even found correlation lacking in two alleles of the same gene. In a related study, Taniguchi *et al.* (*27*) analyzed more than 1000 genes in *E. coli* and measured both mRNA and protein copy numbers in single cells. They found that for most genes, even the numbers of mRNA and protein molecules were uncorrelated. These studies suggest that understanding of regulatory phenomena requires one to consider regulation at both the mRNA and the protein level.

From these studies, it is now clear that variability in single-cell measurements contains a wealth of information that can reveal new insights into the regulatory phenomena of specific genes and the dynamic interplay of entire gene networks. As modern imaging techniques begin to beat the diffraction limitations of light (*28*) and flow cytometers become affordable for nearly any laboratory bench (*29*), we find ourselves in the midst of an explosion in single-cell research. With the advent of single-cell sequencing (*30*, *31*), it might be possible to determine the full transcriptome of many single cells in the near future and to determine the full expression distributions and correlations for all genes in the genome. We expect that the approaches described in this review, which have been pioneered with the model microbial systems, will be readily applied to mammalian cells and tissues (*32*, *33*).

References and Notes

1. G.-W. Li, X. S. Xie, *Nature* **475**, 308 (2011).
2. A. Raj, A. van Oudenaarden, *Cell* **135**, 216 (2008).
3. G. Balázsi, A. van Oudenaarden, J. J. Collins, *Cell* **144**, 910 (2011).
4. A. Eldar, M. B. Elowitz, *Nature* **467**, 167 (2010).
5. M. E. Lidstrom, M. C. Konopka, *Nat. Chem. Biol.* **6**, 705 (2010).
6. B. Snijder, L. Pelkmans, *Nat. Rev. Mol. Cell Biol.* **12**, 119 (2011).
7. D. Zenklusen, D. R. Larson, R. H. Singer, *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).
8. M. Thattai, A. van Oudenaarden, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614 (2001).
9. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, *Science* **280**, 585 (1998).
10. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, *Nat. Methods* **5**, 877 (2008).
11. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, *Cell* **123**, 1025 (2005).
12. J. Peccoud, B. Ycart, *Theor. Popul. Biol.* **48**, 222 (1995).
13. T. B. Kepler, T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
14. J. M. Raser, E. K. O'Shea, *Science* **304**, 1811 (2004).
15. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, *PLoS Biol.* **4**, e309 (2006).
16. V. Shahrezaei, P. S. Swain, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17256 (2008).
17. S. Iyer-Biswas, F. Hayot, C. Jayaprakash, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **79**, 031911 (2009).
18. S. J. Gandhi, D. Zenklusen, T. Lionnet, R. H. Singer, *Nat. Struct. Mol. Biol.* **18**, 27 (2011).
19. L.-H. So *et al.*, *Nat. Genet.* **43**, 554 (2011).
20. R. Z. Tan, A. van Oudenaarden, *Mol. Syst. Biol.* **6**, 358 (2010).
21. S. L. Bumgarner *et al.*, *Mol. Cell* **45**, 470 (2012).
22. L. M. Octavio, K. Gedeon, N. Maheshri, *PLoS Genet.* **5**, e1000673 (2009).
23. J. M. Pedraza, A. van Oudenaarden, *Science* **307**, 1965 (2005).
24. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, *Science* **307**, 1962 (2005).
25. J. Stewart-Ornstein, J. S. Weissman, H. El-Samad, *Mol. Cell* **45**, 483 (2012).
26. M. J. Dunlop, R. S. Cox III, J. H. Levine, R. M. Murray, M. B. Elowitz, *Nat. Genet.* **40**, 1493 (2008).
27. Y. Taniguchi *et al.*, *Science* **329**, 533 (2010).
28. B. Huang, M. Bates, X. Zhuang, *Annu. Rev. Biochem.* **78**, 993 (2009).
29. L. Bonetta, *Nat. Methods* **2**, 785 (2005).
30. T. Kalisky, P. Blainey, S. R. Quake, *Annu. Rev. Genet.* **45**, 431 (2011).
31. F. Tang *et al.*, *Nat. Methods* **6**, 377 (2009).
32. S. Itzkovitz *et al.*, *Nat. Cell Biol.* **14**, 106 (2012).
33. P. Dalerba *et al.*, *Nat. Biotechnol.* **29**, 1120 (2011).
34. B. Munsky, M. Khammash, *J. Chem. Phys.* **124**, 044104 (2006).
35. D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).

Acknowledgments: This work was funded by the National Science Foundation (ECCS-0835623) and a NIH Pioneer award (1DP1OD003936).

10.1126/science.1216379

REVIEW

Computational Approaches to Developmental Patterning

Luis G. Morelli,[1,2,3] Koichiro Uriu,[1,4] Saúl Ares,[2,5,6] Andrew C. Oates[1]*

Computational approaches are breaking new ground in understanding how embryos form. Here, we discuss recent studies that couple precise measurements in the embryo with appropriately matched modeling and computational methods to investigate classic embryonic patterning strategies. We include signaling gradients, activator-inhibitor systems, and coupled oscillators, as well as emerging paradigms such as tissue deformation. Parallel progress in theory and experiment will play an increasingly central role in deciphering developmental patterning.

Animal and plant patterns amaze and perplex scientists and lay people alike. But how are the dynamic and beautiful patterns of developing embryos generated? Used appropriately, theoretical techniques can assist in the understanding of developmental processes (*1*–*5*). There is considerable art in this, and the key to success is an open dialogue between experimentalist and theorist. The first step in this dialogue is to formulate a theoretical description of the process of interest that captures the properties and interactions of the most relevant variables of the system at a level of detail that is both useful and tractable. Once formulated, the second step is to analyze the theoretical model. If the model is sufficiently tractable, it may be possible to understand its behavior with "pencil-and-paper" analysis and compare this analytical solution directly with experimental data. Very often, however, the number of variables and the complexity of their interactions preclude this approach, and the behavior of models must be solved or simulated by using computers in order to be understood and compared with data. This combined approach, which we refer to as computational biology, has become popular recently with the availability of powerful computers and increasingly sophisticated numerical algorithms.

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany. [2]Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, 01187 Dresden, Germany. [3]Consejo Nacional de Investigaciones Científicas y Técnicas, Departamento de Física, Universidad de Buenos Aires, Ciudad Universitaria, 1428 Buenos Aires, Argentina. [4]Theoretical Biology Laboratory, RIKEN Advanced Science Institute, Saitama 351-0198, Japan. [5]Logic of Genomic Systems Laboratory, Centro Nacional de Biotecnología–Consejo Superior de Investigaciones Científicas (CSIC), Calle Darwin 3, 28049 Madrid, Spain. [6]Grupo Interdisciplinar de Sistemas Complejos (GISC), Spain.

*To whom correspondence should be addressed. E-mail: oates@mpi-cbg.de

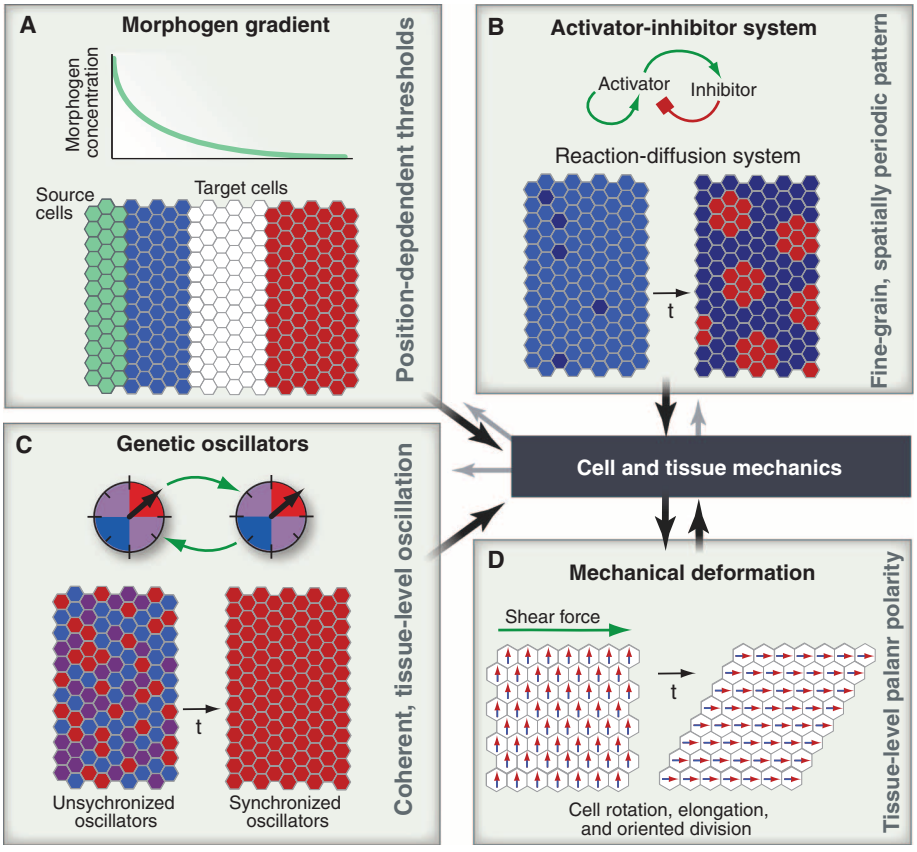

Fig. 1. Patterning strategies. (**A**) Signaling gradients supply global positional information. Horizontal axis is position within target tissue. Morphogen-producing cells are green; cells in tissue take identities (blue, white, and red) according to morphogen concentration. (**B**) Activator-inhibitor systems incorporate local positive and negative feedbacks to generate pattern. Distinct cell types are in red and blue. (**C**) Synchronization of genetic oscillators allows a tissue to generate a coherent temporal rhythm for patterning. In these snapshots, the phase of each oscillating cell is given by its color, which changes over time. (**D**) Tissue deformation can drive patterning reactions. Downstream of patterning information, the dynamic physical properties of tissues drive the morphogenesis of the embryo. *t*, time.

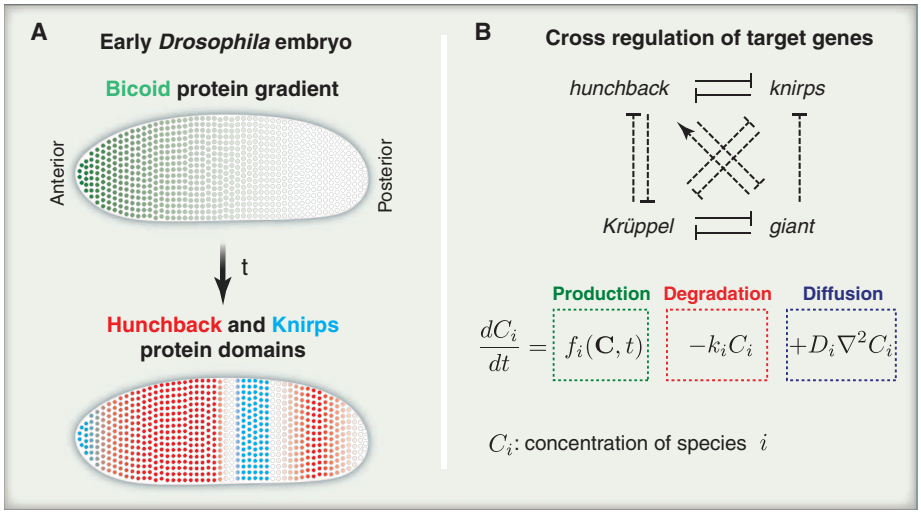

Fig. 2. Patterning with signaling gradients. (**A**) Schematic of early fruit fly embryo showing the maternal gradient of Bicoid protein at cycle 13 that directs the formation of precise target gene domains such as *hunchback* and *knirps*. (**B**) Proposed gene regulatory network showing cross-regulation of target genes (*9*). The four genes are also under control of Bicoid and other players. *t*, time.

In this Review, we hope to introduce scientists familiar with computational methods (geeks) to a selected set of interesting developmental problems (Fig. 1) and to illustrate to developmental biologists (nerds) a selected set of powerful tools. We focus on recent studies investigating four developmental patterning strategies: (i) gradients of signaling molecules released from localized source cells that guide global patterns across target cell populations (Fig. 1A). This external control contrasts with self-organizing strategies within the cell population that use local interactions, such as (ii) activator-inhibitor mechanisms (Fig. 1B) and (iii) the synchronization of cellular oscillations (Fig. 1C). (iv) Mechanical deformations can also change the pattern of a cellular population (Fig. 1D). Although models are often useful in explaining and predicting developmental phenomena, the eventual fate of a given model is to be proven wrong and then modified or replaced, as illustrated in the companion article on cell polarity by Mogilner and colleagues on page 175 of this special issue. Perhaps the greatest impact of computational approaches in developmental biology right now is to force hypotheses to be precisely stated and to stimulate corresponding new quantitative experiments to test them.

Patterning with Signaling Gradients

Morphogens are diffusible signaling molecules that can activate target genes in a concentration-dependent manner. During development, morphogen gradients are established across tissues, diffusing away from localized sources (Fig. 1A). It has been proposed that cells read morphogen levels to determine their position within the tissue and differentiate accordingly (*6*), and there is good evidence that morphogen gradients can direct cell differentiation in target cells. How these gradients are formed, and whether they are sufficient to control differentiation in very precise domains, are open questions that have benefited from computational approaches.

An important model system for studying these questions is the early embryo of the fruit fly *Drosophila*, in part because its geometry and symmetry simplify description and quantitation (Fig. 2A). One of the maternally deposited cues that breaks the symmetry along the embryo's long axis is *bicoid* mRNA, which is present only in the anterior pole. Bicoid protein is translated and transported (*7*), creating within an hour an exponentially decreasing concentration gradient over several hundred micrometers along the embryo's axis. This gradient directs the formation of precise domains of four target genes—among them *hunchback*—that establish the first segments of the future fly body (Fig. 2A). Given the stochastic nature of gene expression, discussed in the companion article by Munsky and colleagues on page 183 of this special issue, morphogen concentration is expected to fluctuate, both over developmental time and from one individual to another. The stunning precision in the position of the boundaries of the segmented out-

put pattern that is found despite these fluctuations puzzles both nerds and geeks. The field has wrestled with the issue of whether this precision can be achieved through the Bicoid gradient alone, or whether other mechanisms are required.

Contributing to this debate, recent papers by Manu *et al*. (*8*, *9*) formulated the interactions between four target genes downstream of the maternal gradients in the early embryo using a gene regulatory network (GRN) model, in which each variable represents the quantity of a molecular species (Fig. 2B). One of the limitations of GRN models is that great experimental effort is often required to estimate relevant values of the model's many parameters in the embryo. Parameters for this *Drosophila* segmentation model were obtained computationally by finding those combinations that best reproduced a time series of quantitative spatial gene expression data from the embryo. The model hinted that cross-regulatory interactions between target genes in the GRN reduce the variability in the position of their expression domains.

One problem in understanding a model is that as the parameters vary, the general dynamic behavior of the system can change dramatically. These changes are called bifurcations, and using powerful tools from dynamical systems theory (*10*), Manu *et al*. (*9*) performed a bifurcation analysis of the model to identify the fundamental behaviors that the system can display over a given set of realistic parameter values. The model predicts that cells in the anterior of the embryo select a stable state of the dynamics, and the concentrations of targets change as Bicoid levels drop. In the posterior of the embryo, the system never reaches a stable state because gastrulation happens first. Describing the simple behaviors of a complex regulatory network in this compact way is appealing because it makes similarities to other regulatory systems clearer and also makes falsifiable predictions about distinctive behaviors that can be experimentally tested.

Fluctuations in gene product levels generate molecular noise that limits the precision of signaling gradients and also degrades the targets' outputs. This problem can be formulated precisely by using the tools and concepts from information theory—originally used in engineering—which quantifies the flow of information through communication channels. A key concept is the mutual information between two variables, such as, for example, Bicoid and Hunchback levels. An elegant computation by Tkačik and Walczak used existing precise measurements of morphogen levels (*11*) to estimate the mutual information between Bicoid and Hunchback (*12*). On the basis of their result, they argued that if similar results hold for the other target genes under Bicoid control, the combined information conveyed by the four genes would be enough so that each of the roughly 100 rows of nuclei could unambiguously determine its position along the *Drosophila* embryo. To test this hypothesis, combined high-quality spatial expression data for the other target genes in the system will be necessary. Thus, information theory is emerging as a potentially powerful tool to quantify information transmission in developmental GRNs. As yet, it is unclear whether the *bicoid* gradient is sufficiently precise to instruct the precise boundaries of its target gene domains, or whether other mechanisms are necessary, but computational biology has a central role in this discussion.

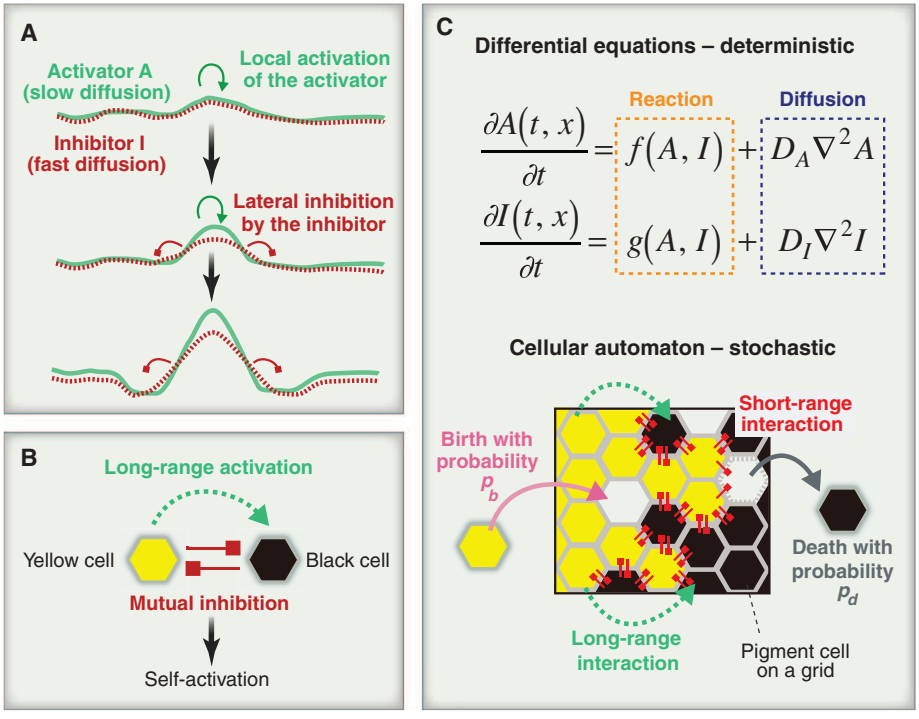

Fig. 3. Patterning with activator-inhibitor systems. (**A**) Local activation and lateral inhibition generates spatially heterogeneous patterns. (**B**) Interactions between black and yellow pigment cells produce Turing patterns in zebrafish skin. Mutual inhibition between them functions as self-activation for the yellow cells. Each yellow cell activates distant black cells. Therefore, inhibition of the yellow cell by the black cell works as a lateral inhibition. (**C**) Different modeling approaches to spontaneous pattern formation.

Patterning with Activator-Inhibitor Systems

Cells in a morphogen gradient use the local level of an externally provided signal to produce patterns (Fig. 1A). However, patterns such as spots and stripes can arise spontaneously from entirely local interactions. In 1952, Alan Turing proposed a reaction-diffusion (RD) mechanism to explain spontaneous pattern formation without signaling gradients (*13*). Specifically, he considered two diffusing chemical components, an activator and an inhibitor (Figs. 1B and 3A). By self-activation, the activator can locally increase its concentration (Fig. 3A). The activator in that region produces the inhibitor, which suppresses the activator in surrounding space because of faster diffusion. As a result, local peaks of activator self-organize from the almost homogeneous starting state, leading to the spontaneous formation of spatial patterns, such as stripes and spots in a two-dimensional (2D) space (so-called Turing patterns) (Fig. 1B).

Subsequently, RD systems have been considered to play important roles in spontaneous pattern formation (*14*, *15*). Although spatial structures very similar to simulated Turing patterns have been observed in development, until recently there was scant evidence showing that the Turing mechanism causes these structures. Indeed, conceptually elegant RD models of the *Drosophila* segmentation process introduced above proved to be entirely wrong (*16*), and this failure may even have left some developmental biologists wary of further theoretical efforts. However, identification of interaction rules and key molecular components in several putative RD systems (*17*, *18*) now suggests the potential of a long-awaited experimental verification of these ideas.

Skin pattern formation in fish has long been a candidate for patterning by use of the Turing mechanism (*19*). To identify key interaction rules in the system, Nakamasu *et al*. studied stripe formation in zebrafish skin (*20*). These black and yellow stripes are self-organized over 3 weeks by local interactions between black and yellow pigment cells, which fulfill the condition for Turing patterns (Fig. 3B). To confirm that the experimentally observed interactions between pigment cells can generate stripes, the authors first used deterministic partial differential equations to model cellular dynamics. However, because the width of each stripe in zebrafish is only ~10 cells, Nakamasu *et al*. pointed out that stochastic effects caused by smaller cell numbers might prevent stable stripe formation. In that situation, it would

be a better formulation to explicitly describe stochastic behaviors of each single pigment cell, such as birth, movement, and cell death. The authors developed a cellular automaton-based model (Fig. 3C) that includes the observed pigment cell interactions to study the robustness of stripe patterns against stochastic effects. Although such detailed models usually include several parameters not measured experimentally, simulations of the cell-based model produced patterns similar to those obtained by the deterministic model and observed on the zebrafish skin. Combining investigations of the molecular and cellular basis of the cellular-level interaction rules (*21*) with further theoretical studies should reveal whether this is indeed a Turing system.

Gradient patterning strategies can also be formulated as RD systems because gradients can arise from diffusion of morphogens, and the pattern emerges due to reactions that involve these morphogens. However, the different length-scales involved in activator-inhibitor systems give rise to qualitatively different patterns, which are local in nature. This is an example of how very different developmental patterning strategies can be described by using similar model formulations.

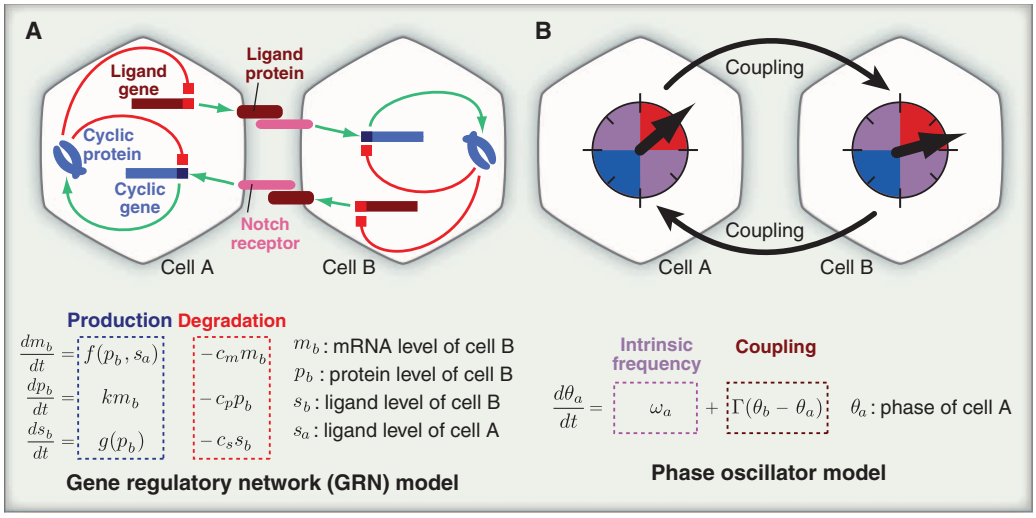

Fig. 4. Patterning with genetic oscillations. (**A**) Cyclic gene expression oscillates in individual cells because of a negative feedback loop, and oscillations are coupled to neighbor cells through the Notch pathway. (**B**) The mutual effects of cellular oscillators can be described by models of coupled phase oscillators.

Patterning with Genetic Oscillations

The growing body axis of all vertebrate embryos is rhythmically and sequentially subdivided into segments. For example, in the zebrafish embryo the multicellular segments are ~50 μm long and form with a periodicity of 30 min. Inspired by such clock-like regularity, Cooke and Zeeman proposed the Clock and Wavefront model in 1976 (*22*). In this model, a biological clock ticks at the posterior of the elongating embryo, and the distance advanced by a wavefront along the embryonic axis during a cycle of the clock sets the length of a forming segment. More than 20 years later, the model was revived with the discovery of genetic oscillations in the chick embryo (*23*). This segmentation clock appears to be a tissue-level rhythmic pattern generator (*24*), in which a population of progenitor cells behave as coupled oscillators, self-organizing a collective rhythm through mutual synchronization (Fig. 1C).

A clue to the existence of such a synchronized cell population came from zebrafish mutants that disrupt Delta-Notch intercellular signaling, in which coherent oscillations and segmental patterning are gradually lost (*25*). The current hypothesis is that in the wild-type embryo, Delta ligands under the control of a single-cell oscillator activate Notch receptors in the membrane of neighboring cells, and these receptors coordinate oscillating gene expression in the receiving cell (Fig. 4A). Without Delta-Notch signaling, the single cells' oscillations gradually lose synchrony. The plausibility of this synchronization hypothesis has been studied by using GRN models showing that the Delta-Notch mechanism described above could keep neighboring cells oscillating in synchrony (*26*, *27*).

Given the previously mentioned difficulty of determining GRN parameters from embryos (*28*), an alternative and complementary model formulation is to use an effective theory with variables that represent processes for which there is a particular interest or a possibility of experimental comparison. For the segmentation clock, this approach has been applied to investigate the synchronization hypothesis by using theories based on coupled phase oscillators (Fig. 4B). In a phase oscillator model, the variables corresponding to oscillating molecular species are substituted by a single variable: the phase of the oscillation cycle, which advances in time with a given intrinsic frequency. The effect of Delta-Notch signaling is captured by a coupling function that speeds up or slows down a cellular oscillator depending on the phase of neighboring cells. Phase oscillator models do not offer direct insight about dynamics of individual molecular species, but their simplicity allows powerful insights about system-level dynamics from paper-and-pencil analysis. Furthermore, they allow a direct fit to experimental data relying on a few coarse-grained parameters such as the period of the oscillations (*29*).

Using a phase oscillator model, the synchronization problem of the segmentation clock was formulated as a competition between noise and the intercellular coupling that keeps cells in synchrony (*30*). Together with quantitative experimental disruptions of Notch signaling in zebrafish, the model allowed estimation of the noise level and coupling strength relevant for the tissue-level synchrony of the clock. Coupling involves the new synthesis of Delta ligand every cycle (Fig. 4A), and to represent the anticipated duration of the ligand-receptor mechanism, Morelli *et al.* (*29*) included explicit time delays in the coupling function of a phase oscillator model. This delayed coupling theory made the prediction that changing the coupling strength could change the clock period and motivated the study of the dynamics of Notch mutants. Quantitative time-lapse measurements of segmentation period and analysis of clock gene-expression patterns in mutants matched the theoretical predictions and so identified the first candidates for segmentation clock period mutants (*31*).

Although these studies have revealed some surprising insights into the segmentation clock's dynamics, most quantitative data used to test models have come from static images (*28*, *31*), and the desynchronization of the clock has not been directly observed. The advent of new techniques to observe cyclic gene expression in vivo (*32*) will allow key assumptions of the existing models to be directly tested.

Patterning with Mechanical Deformations

We complete our roster of patterning mechanisms with a recently discovered case driven by tissue deformations. An apparently simple behavior for an epithelial sheet is to elongate along one axis while shrinking along the orthogonal axis. During *Drosophila* development, the wing blade epithelium stretches into the familiar elongate wing shape, and each of the hairs protruding from the wing cells points distally—an example of planar cell polarity (PCP) patterning (Fig. 5A). Although proximo-distal gradients of PCP pathway components have been observed, they are not sufficient to produce the final wing hair polarity (*33*). Examination of cell shapes and trajectories from time-lapse movies shows that sharp contraction of the neighboring hinge region exerts anisotropic tension on the wing blade (*34*). Over a period of 15 hours, the blade deforms with a shear gradient arising from the cellular flow in the tissue.

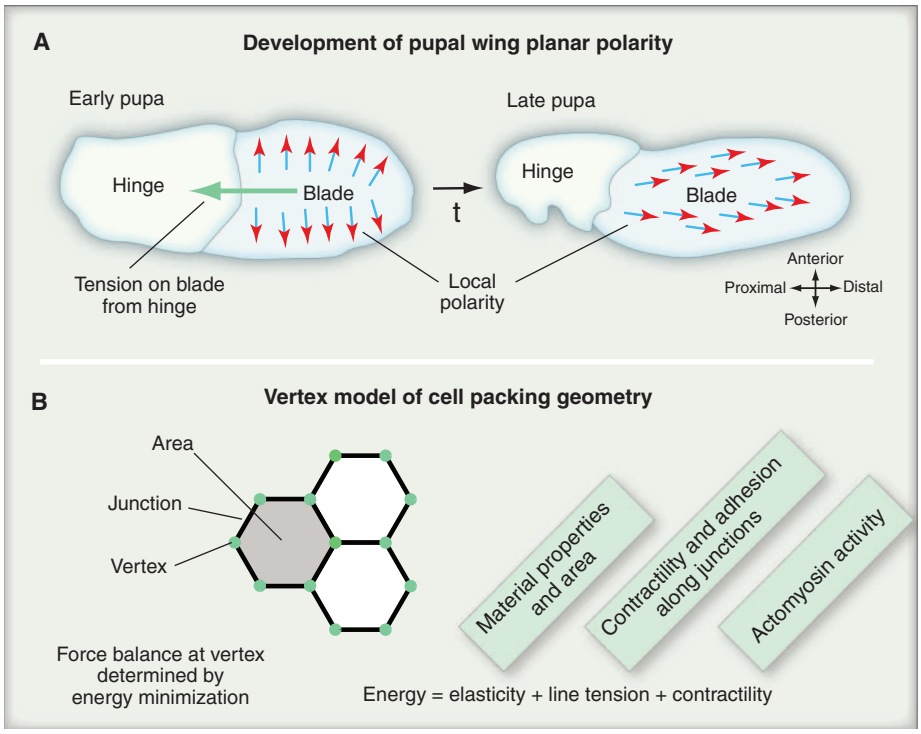

Fig. 5. Patterning by mechanical deformation. (**A**) Overview of *Drosophila* wing development during pupariation, when the wing blade elongates and proximo-distal planar polarity is established. (**B**) Schematic of the vertex model used to calculate stable cell-packing geometries.

Aigouy *et al*. explored the role of tissue shear in aligning the axis of cellular polarity with the proximo-distal axis of the wing blade by formulating a 2D vertex model of epithelial cell shape (Fig. 5B) (*35*), incorporating an effective description of the local recruitment of complementary PCP molecules to apposing cell boundaries (*34*). This new model predicts that polarity is reoriented by local rotation and cell flow–induced shear. Simulations show that shear associated with oriented cell division, proximo-distal cell elongation, and cell rearrangement also contribute to the alignment of cell polarity with the long axis of the wing. Future work can investigate how the 3D baso-lateral surfaces of the epithelial cells in the wing affect this description, and how the PCP protein complexes involved dynamically reorganize during cellular rearrangement. Thus, remarkably the final planar cell polarity of the completed wing may be a direct consequence of the externally applied stresses responsible for its extension, via simple physical rules such as those that determine molecular polarity in liquid crystals (*36*).

In this Review, we have mainly discussed chemical aspects of pattern formation as separate from downstream mechanics of morphogenesis (*37*, *38*). Turing already wondered whether a closer linkage might be at work (*13*), and it seems timely to reconsider development as having integrated mechanochemical aspects (*39*). For example, motivated by recent findings on cell cortex dynamics in the nematode *Caenorhabditis* (*40*), Bois *et al*. studied pattern formation in an active fluid in which mechanical contraction causes the flow of reactive chemical species (*41*). This theoretical analysis showed that an active fluid extends the parameter space in which classical Turing systems generate spatial patterns. To what extent continuous feedback between chemical and mechanical processes also underlies tissue-level phenomena in development is not yet clear, but it may be widespread.

Outlook

With the wide range of approaches in use, how should the developmental biologist select the appropriate modeling and computational methods? And where should the computational scientist dig for interesting problems in the vast field of developmental biology? Previous reviews have given multiple examples and advice (*1*–*5*). Here, we argue that the first step is key: The level of description and model type should be matched to the best available data. The data should be quantitative, accurate, and precise, and the model should make falsifiable predictions. Although some researchers are fluent in both domains, most often a successful computational approach to developmental biology will involve a long-term dialogue between experts across disciplinary boundaries. As advances in imaging and molecular methods increase experimental resolution and complexity, corresponding theoretical and computational developments will be required to assemble the puzzle. This co-dependence should generate a wealth of new opportunities for geeks and nerds alike.

References and Notes

1. A. C. Oates, N. Gorfinkiel, M. González-Gaitán, C. P. Heisenberg, *Nat. Rev. Genet.* **10**, 517 (2009).
2. J. Lewis, *Science* **322**, 399 (2008).
3. G. T. Reeves, C. B. Muratov, T. Schüpbach, S. Y. Shvartsman, *Dev. Cell* **11**, 289 (2006).
4. C. J. Tomlin, J. D. Axelrod, *Nat. Rev. Genet.* **8**, 331 (2007).
5. S. Roth, *Dev. Genes Evol.* **221**, 255 (2011).
6. L. Wolpert, *J. Theor. Biol.* **25**, 1 (1969).
7. S. C. Little, G. Tkačik, T. B. Kneeland, E. F. Wieschaus, T. Gregor, *PLoS Biol.* **9**, e1000596 (2011).
8. Manu *et al*., *PLoS Biol.* **7**, e1000049 (2009).
9. Manu *et al*., *PLOS Comput. Biol.* **5**, e1000303 (2009).
10. S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Westview, Boulder, CO, 1994).
11. T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, D. W. Tank, *Cell* **130**, 141 (2007).
12. G. Tkačik, A. M. Walczak, *J. Phys. Condens. Matter* **23**, 153102 (2011).
13. A. M. Turing, *Philos. Trans. R. Soc. London Ser. B* **237**, 37 (1952).
14. H. Meinhardt, *Models of Biological Pattern Formation* (Academic Press, London, 1982).
15. J. D. Murray, *Mathematical Biology* (Springer Verlag, Berlin, 2003).
16. M. Akam, *Nature* **341**, 282 (1989).
17. A. D. Economou *et al*., *Nat. Genet.* **44**, 348 (2012).
18. M. V. Plikus *et al*., *Science* **332**, 586 (2011).
19. A. Kondo, R. Asai, *Nature* **376**, 765 (1995).
20. A. Nakamasu, G. Takahashi, A. Kanbe, S. Kondo, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8429 (2009).
21. M. Inaba, H. Yamanaka, S. Kondo, *Science* **335**, 677 (2012).
22. J. Cooke, E. C. Zeeman, *J. Theor. Biol.* **58**, 455 (1976).
23. I. Palmeirim, D. Henrique, D. Ish-Horowicz, O. Pourquié, *Cell* **91**, 639 (1997).
24. A. C. Oates, L. G. Morelli, S. Ares, *Development* **139**, 625 (2012).
25. Y. J. Jiang *et al*., *Nature* **408**, 475 (2000).
26. J. Lewis, *Curr. Biol.* **13**, 1398 (2003).
27. K. Uriu, Y. Morishita, Y. Iwasa, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4979 (2010).
28. F. Giudicelli, E. M. Ozbudak, G. J. Wright, J. Lewis, *PLoS Biol.* **5**, e150 (2007).
29. L. G. Morelli *et al*., *HFSP J.* **3**, 55 (2009).
30. I. H. Riedel-Kruse, C. Müller, A. C. Oates, *Science* **317**, 1911 (2007).
31. L. Herrgen *et al*., *Curr. Biol.* **20**, 1244 (2010).
32. Y. Niwa *et al*., *Genes Dev.* **25**, 1115 (2011).
33. D. Ma *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18800 (2008).
34. B. Aigouy *et al*., *Cell* **142**, 773 (2010).
35. R. Farhadifar, J. C. Röper, B. Aigouy, S. Eaton, F. Jülicher, *Curr. Biol.* **17**, 2095 (2007).
36. J. F. Joanny, F. Julicher, K. Kruse, J. Prost, *New J. Phys.* **9**, 422 (2007).
37. N. Gorfinkiel, S. Schamberg, G. B. Blanchard, *Genesis* **49**, 522 (2011).
38. S. M. Trier, L. A. Davidson, *Curr. Opin. Genet. Dev.* **21**, 664 (2011).
39. J. Howard, S. W. Grill, J. S. Bois, *Nat. Rev. Mol. Cell Biol.* **12**, 392 (2011).
40. M. Mayer, M. Depken, J. S. Bois, F. Jülicher, S. W. Grill, *Nature* **467**, 617 (2010).
41. J. S. Bois, F. Jülicher, S. W. Grill, *Phys. Rev. Lett.* **106**, 028103 (2011).

Acknowledgments: We thank F. Jülicher, C.-P. Heisenberg, S. Grill, P. R. ten Wolde, T. Bollenbach, P. Formosa-Jordan, and members of the Oates and Jülicher groups for discussion and critical comments. This work was supported by the Max Planck Society and the European Research Council (ERC) under the European Communities Seventh Framework Programme (FP7/2007-2013)/ERC grant 207634. K.U. is supported by the Japan Society for the Promotion of Science for Young Scientists. S.A. acknowledges funding from CSIC through the "Junta para la Ampliación de Estudios" program (JAEDOC014, 2010 call) co-funded by the European Social Fund, and from Ministerio de Ciencia e Innovación (Spain) through MOSAICO.

10.1126/science.1215478

PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Opinion piece

Cite this article: Sarma GP *et al*. 2018 OpenWorm: overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. B* **373**: 20170382.
http://dx.doi.org/10.1098/rstb.2017.0382

Accepted: 16 August 2018

One contribution of 15 to a discussion meeting issue 'Connectome to behaviour: modelling *C. elegans* at cellular resolution'.

Subject Areas:
behaviour, biophysics, neuroscience, physiology, systems biology, theoretical biology

Keywords:
computational neuroscience, bioinformatics, software engineering, computational physiology, biological simulation, *Caenorhabditis elegans*

Author for correspondence:
Stephen D. Larson
e-mail: stephen@openworm.org

OpenWorm: overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*

Gopal P. Sarma[1], Chee Wai Lee[2], Tom Portegys[3], Vahid Ghayoomie[4], Travis Jacobs[5], Bradly Alicea[6], Matteo Cantarelli[2], Michael Currie[7,14], Richard C. Gerkin[16], Shane Gingell[15], Padraig Gleeson[10], Richard Gordon[11,12], Ramin M. Hasani[13], Giovanni Idili[2], Sergey Khayrulin[2,8,9], David Lung[13], Andrey Palyanov[8,9], Mark Watts[14] and Stephen D. Larson[2]

[1]School of Medicine, Emory University, Atlanta, GA, USA
[2]The OpenWorm Foundation, New York, NY, USA
[3]Ernst and Young LLP, New York, NY, USA
[4]Laboratory of Systems Biology and Bioinformatics, University of Tehran, Tehran, Iran
[5]Melonport AG, Zug, Switzerland
[6]Orthogonal Research, Champaign, IL, USA
[7]Fling Inc., Bangkok, Thailand
[8]Laboratory of Complex Systems Simulation, A.P. Ershov Institute of Informatics Systems, Novosibirsk, Russia
[9]Laboratory of Structural Bioinformatics and Molecular Modeling, Novosibirsk State University, Novosibirsk, Russia
[10]Department of Neuroscience, Physiology and Pharmacology, University College London, London, UK
[11]Embryogenesis Center, Gulf Specimen Marine Laboratory, Panacea, FL, USA
[12]C.S. Mott Center for Human Growth and Development, Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI, USA
[13]Cyber-Physical Systems, Technische Universität Wien, Wien, Austria
[14]Raytheon Company, Waltham, MA, USA
[15]Out of the BOTS, Inc., Queensland, Australia
[16]School of Life Sciences, Arizona State University, Tempe, AZ, USA

GPS, 0000-0002-9413-6202; BA, 0000-0003-3869-3175; RCG, 0000-0002-2940-3378; PG, 0000-0001-5963-8576; AP, 0000-0003-1108-1486; MW, 0000-0002-6782-8510; SDL, 0000-0001-5397-6208

The adoption of powerful software tools and computational methods from the software industry by the scientific research community has resulted in a renewed interest in integrative, large-scale biological simulations. These typically involve the development of computational platforms to combine diverse, process-specific models into a coherent whole. The OpenWorm Foundation is an independent research organization working towards an integrative simulation of the nematode *Caenorhabditis elegans*, with the aim of providing a powerful new tool to understand how the organism's behaviour arises from its fundamental biology. In this perspective, we give an overview of the history and philosophy of OpenWorm, descriptions of the constituent sub-projects and corresponding open-science management practices, and discuss current achievements of the project and future directions.

This article is part of a discussion meeting issue 'Connectome to behaviour: modelling *C. elegans* at cellular resolution'.

1. Introduction

In 2011, the OpenWorm project was launched with the mission of building the world's first detailed biophysical simulation of the nematode *Caenorhabditis elegans* [1,2]. In addition to the ambitious scientific goals, a unique aspect of the project is the fully open science, distributed research framework in which the work would take place. In this article, we look at the past, present and future of OpenWorm. What has it achieved in the period since its foundation,

THE ROYAL SOCIETY PUBLISHING

what are the important next steps and what can others learn from this experience?

A unifying principle underpinning OpenWorm is the application of an engineering approach to the challenge of managing biological complexity [3]. Modern software engineering has given us the tools to keep track of the hundreds of thousands of details of which complex physical systems are composed. The synergy between human and machine in computer-assisted modelling can allow for deeper reasoning than either a human or computer alone. In industrial manufacturing, for example, advances in engineering software have enabled materials simulations that allow mechanical engineers to test many different mechanisms *in silico* before the manufacturing process [4]. While the fields of computational biology and computational neuroscience have made significant advances over their multi-decade history, simulations have only had a limited impact on the biological thinking process when compared with other disciplines in the physical and engineering sciences [5].

What level of complexity should our model aim for? Our perspective is the following: *an integrative model need not incorporate any more detail than the individual models the research community has already produced*. In other words, we take a holistic approach in which individual models, which may operate at multiple scales, are thoughtfully integrated into a unified computational platform, providing a global view of the entire organism. As we will discuss below, the lowest level of biological detail that the OpenWorm project incorporates is that of ion channel models which underpin membrane potential dynamics. Examples such as the whole cell model of Karr *et al.* [6] demonstrate that this kind of 'holistic biology' can lead to valuable insights into underlying biological function [6]. The purpose of this work is integrative and allows us to extract even greater value from the knowledge the scientific community has already produced. Nowhere is the need for models that encompass multiple scales more evident than in the hermaphrodite nematode's network of 302 neurons, where simple crawling and swimming behaviours remain unexplained [7]. Despite decades of effort, we struggle to describe how individual neurons give rise to such diverse organismal behaviour. Our belief is that a computational platform in which an organism's behaviour arises from lower-level biological models will come to play a significant role in advancing the field.

In the field of *C. elegans* biology, there has been significant effort to collect comprehensive anatomical and other structural data about the nervous system, ranging from the electrical and synaptic connectome [8], to cholinergic and GABAergic neurons [9,10], to the extrasynaptic connectome of neuropeptides [11]. The purpose of generating these 'map'-like datasets is to communicate the relationships between biological entities. Unfortunately, the complexity of such datasets places severe limitations on their intelligibility. This problem is not unrelated to modern genomics, where the many-tangled webs of relationships between hundreds of thousands of genes and gene products demand computational tools to assist in their understanding. It is with this complexity in mind that the OpenWorm project has taken upon itself to integrate the disparate and heterogeneous physiological maps and related datasets generated by the *C. elegans* community into a coherent software framework. Efforts such as PyOpenWorm (described below) are one such example in OpenWorm where publicly available data are assembled into a graph database and Python application programming interface, enabling users to query multiple datasets about *C. elegans* neuronal structure. By creating an open, shared repository and query tool for these data, the fruits of collective labour become integrated into a shared structure that amplifies the impact of the entire community's research output. Moreover, the need to arrive at a global view of relevant datasets has allowed us to identify key areas where new data should be collected, potentially taking advantage of novel experimental apparatus such as robotic patch-clamp set-ups. Ultimately, we expect that unified platforms for data integration will dovetail with other contemporary efforts in the life sciences to increase the robustness and exchangeability of datasets and models [12–14].

In the scientific community, assembling datasets solely for the purpose of consolidation has often led to the emergence of multiple, redundant standards. In the OpenWorm project, our fundamental aim is to curate datasets and mathematical models in a manner that facilitates dynamic simulations of biological function. Theoretical biophysicists have produced a rich literature of quantitative models of *C. elegans* physiology, ranging from membrane potential dynamics, to neuromuscular coupling, to the fluid dynamics of body movement. Integrating these individual models into a global, composite simulation creates an additional check on the underlying datasets themselves. In addition, the simulation enables the construction of complex hypotheses which researchers can further investigate through theoretical or experimental means [5].

In deciding on the level of biological detail we wish to incorporate, we have agreed upon an approach that incorporates biomechanics as a critical component of understanding the nematode in the context of its environment [15]. In addition to the biological implications, maintaining biological realism may have implications well beyond understanding *C. elegans*. Indeed, researchers in the artificial intelligence community have posited that sensorimotor feedback may play a role in allowing future AI systems to learn from experience more efficiently than current data-hungry systems based on deep learning [16]. As such, we have unified a biomechanical model of *C. elegans*, Sibernetic [17,18], that incorporates interactions with a fluid or gel environment, with a modelling infrastructure for complex neuronal networks, c302 [19].

Our ultimate vision for OpenWorm is to provide a computational platform that allows for simulations to become seamlessly integrated into biological thought. Rather than replacing existing theoretical or experimental methods, our vision is to take advantage of the powerful tools of modern software engineering to maximally enable the research community and leverage long-standing intellectual traditions and biological insights [5]. We can imagine a number of possible applications for such a platform. Because we have complete control over all details of the simulation, we can effortlessly create knockouts, where, for example, all synaptic connections to or from a specific cell can be removed. We can simulate known mutants that have ion channels with different properties and observe their behaviour. We can simulate the effects of drugs by modelling their impact on ion channels, potentially paving the way to using simulations as a way to generate hypotheses for new uses of existing pharmacological agents and for discovering new ones. If successful in the *C. elegans* community, we would hope this approach could assist in the understanding of

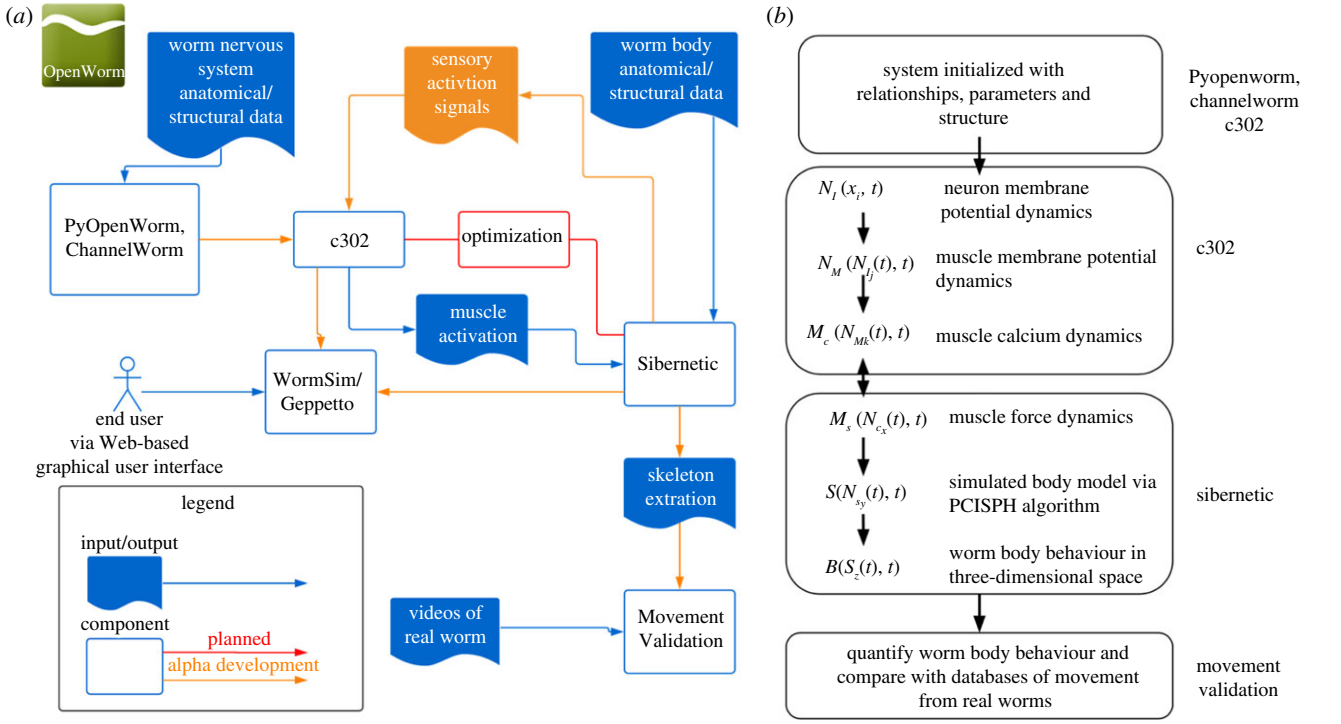

Figure 1. Overview of OpenWorm simulation stack. (*a*) A component diagram describing the relationships between inputs and outputs of sub-projects within OpenWorm. (*b*) A highly simplified schematic view of the system of equations executed in the combined c302/Sibernetic system.

other organisms in biology. In the remainder of this paper, we describe progress in the open resources we have produced, their uses and features, and future directions for the project.

2. Material and methods

(a) Software infrastructure for simulating *C. elegans*

(i) OpenWorm simulation stack

OpenWorm is organized into a number of sub-projects, several of which are described in more detail in this issue. In this section, we will give a condensed overview of the core of the platform. A 'simulation stack' refers to the set of integrated software tools that are used to run a simulation. It is called a 'stack' because each tool can be thought of as existing at a certain level in a hierarchy of abstraction and information flow. For instance, at the lowest level, we have ion channel models and connectomes. At the next level, we have models of neuromuscular coupling. And finally, the output generated by the connectome can be fed into a simulation of the body movement and environment. While each element of this simulation stack could form the basis for an independent research project, our aim is to use best practices from the software industry to integrate these tools into a single software framework. Figure 1 shows the different components of the OpenWorm simulation stack and their relationships.

Figure 1*a* shows a breakdown of the contents of the OpenWorm simulation stack described as components of software. Inputs and outputs to the software components are depicted with arrows showing how they relate to the core modules. For the PyOpenWorm and ChannelWorm software projects, inputs include anatomical and structural data from the worm's nervous system and knowledge about ion channels, respectively. These data are fed into the c302 software component, which constructs systems of equations that are used to simulate the membrane dynamics of the nervous system at multiple levels of detail ranging from simple integrate-and-fire neurons to multi-compartment neuronal models [19]. The outputs of the c302 simulation include muscle activation signals which form the inputs to the Sibernetic system. We are planning on incorporating feedback from Sibernetic to c302 that represents sensory signals generated from the worm body's posture as well as interactions with the environment. Additionally, Sibernetic takes as an input structural and biophysical data about the worm. The output from Sibernetic, the outline of the worm's body as it bends and moves over time, can be fed into the movement validation software system, where comparisons with videos of real worms are used to validate the global model's biological validity. These two systems will be incorporated into a web-based graphical user interface framework that provides a visual interface to the end user via WormSim/Geppetto [20]. An optimization block in the diagram indicates where the free parameters in the models can be filled in by tuning model parameters of single neurons to match experimental data [21,22].

In figure 1*b*, we show a simplified schematic that breaks down the integrated c302/Sibernetic system into mathematical components. The system is initialized with relationships, parameters and structure derived from databases that have been populated with information about *C. elegans* physiology. In the current version of the simulation, we begin with neuron membrane potential dynamics (N_I) that are set manually. From those dynamics, the electrical activities of the body wall muscle cells (N_M), i.e., those muscles receiving direct synaptic input from neurons, are calculated. This activation also results in dynamical changes in the muscles' internal calcium concentration (M_c). These first components of the simulation are carried out in the c302 framework, which executes a NeuroML-based model in the NEURON simulation engine [19]. The calcium dynamics of the muscle cells calculated by c302 (M_c) are passed into Sibernetic as activation signals. These activation signals are converted into forces that cause activated muscle cells lining the body model to contract (M_s). The combination of the contraction states of all the muscles leads to the state of the simulated body model as a whole (S), calculated via the predictive–corrective incompressible smoothed particle hydrodynamics (PCISPH) algorithm for

modelling fluids. The aggregate of all of the particles that make up the simulated body model is the behaviour of the simulated worm over time (B). This can then be compared against the movement of real worms once brought into a comparable format [23]. While there is currently only a uni-directional flow of information from c302 to Sibernetic via muscle activation signals, we are developing a reverse step where forces on the skin of the worm body model lead to activation signals of sensory neurons.

(ii) PyOpenWorm

Biological data are often weakly structured and heterogeneous, which creates fundamental problems for computational platforms that rely on these data. In addition, discrepancies that are frequently seen between database formats and term definitions create even further difficulties for end users. The challenges in making use of biological data are common across all subfields of computational biology, with *C. elegans* being no exception. PyOpenWorm (https://github.com/openworm/pyopenworm) is a Python package intended to simplify access to a range of structured data on *C. elegans* anatomy and physiology. It is a data access layer for *C. elegans* information, where users can query data across multiple scales of the worm's biology. The heterogeneous nature of *C. elegans* biology requires that different underlying technologies be used to store different types of data. For instance, an RDF semantic graph representation is useful for representing neuronal structural properties such as ion channel expression and the density and type of neurotransmitter receptors, whereas a NeuroML representation is most appropriate for storing model morphology and simulation parameters [24]. PyOpenWorm solves the problem of abstracting away the underlying technologies, so the user can query the system in a manner that is intuitive for researchers who are already familiar with the worm's biology. The resulting data can be used directly or as part of a multistage software pipeline. The software project is implemented in the Python programming language and the code is available on GitHub along with the other sub-projects of OpenWorm.

Data from reliable external sources, most often published journal articles, are collected into a single directory in the PyOpenWorm repository. These datasets can take the form of structured spreadsheet files or even other relational databases. For quality control, we only consider data that have an original source associated with them. Currently, data are collected from the literature and other secondary sources, such as WormBase [25] and WormAtlas [26]. When a user or program connects with PyOpenWorm's database, they have access to all of the data through a simple Python library. Table 1 lists current data sources that are part of PyOpenWorm.

Although PyOpenWorm's primary current use cases are for storing static data and models, its fundamental architecture anticipates future needs once members of the research community begin to make use of the OpenWorm tool stack as part of their daily research. In particular, ongoing development of PyOpenWorm is aimed at ensuring that the system can store metadata and simulation results, so that this output can subsequently be interrogated and analysed as part of the research process.

(iii) ChannelWorm

As we discussed above, ion channels represent the most granular level of biological detail that the OpenWorm simulation incorporates. Ion channels are pore-forming proteins, found in the membranes of all cells. They are responsible for many known cellular functions including shaping action potentials and gating the flow of ions across the cell membrane. Remarkably, most nematode ion channels are conserved across vertebrate species [28]. Because of their widespread relevance for biology, many electrophysiological experiments have been focused on ion channels and transporter functional genomics in *C. elegans* [29–34]. Although much of this work is experimental, computational work has also been directed at integrating ion channel models into larger-scale simulations [35–40]. One such example (outside *C. elegans* biology) is the Blue Brain Project, which recently unveiled a detailed simulation of a rat cortical microcolumn [41], taking advantage of an extensive repository of curated data and models of ion channels [42].

We have chosen ion channel models as the most granular level of detail with which to simulate the nematode for several reasons. For instance, insights into drug development would not be possible without an understanding of the action of the major neurotransmitter species on Na^+, K^+ and Ca^{2+} currents. Fortunately, incorporating ion channel models is a tractable approach and there is no need to limit ourselves to simulations of more abstract neurons. Moreover, the specific dynamics of ion channels themselves are key components of the models of neuromuscular coupling that we use. And as we argued above, biomechanics is a central component of our scientific roadmap.

As part of the OpenWorm project, we created ChannelWorm (https://github.com/openworm/channelworm and https://chopen.herokuapp.com) in order to (i) integrate and structure data related to ion channels in *C. elegans*, (ii) digitize and curate electrophysiological data from publications, (iii) develop application programming interfaces for accessing these data and (iv) build ion channel models based on experimental data. As the project has progressed, we have found ourselves in the unique position of attempting to develop a global view of the current state of *C. elegans* ion channel modelling. One of the major lessons we have learned is that patch-clamp data are only available for a small minority of ion channels expressed in the nematode. Consequently, a significant undertaking is to build Hodgkin–Huxley models for ion channels that lack these data based on homologous channel types from other organisms. After a manual curation process in which contributors digitize electrophysiological plots, kinetic parameters are derived from these data using genetic algorithms [22,43,44] and related techniques such as particle swarm optimization. Ultimately, these ion channel models are translated into the NeuroML markup language [45,46], which allows for consistent representation of neuronal biophysics, anatomy and network architecture for use in subsequent computational simulations.

Table 1. Data sources incorporated into PyOpenWorm (https://pyopenworm.readthedocs.io/en/latest/data_sources.html).

data type	source
neurons and muscles	
names	WormBase; Harris *et al*. [25]
neuron types, cell descriptions, lineage names, neurotransmitters, neuropeptides, receptors, innexins	WormAtlas; Altun *et al*. [27]
monoamine secretors and receptors, neuropeptide secretors and receptors	Bentley *et al*. [11]
connectome	
neuron to neuron and neuron to muscle chemical synapses and gap junctions	personal communication by S. Cook (original data set available at http://bit.ly/2MGiv9K); White *et al*. [8]

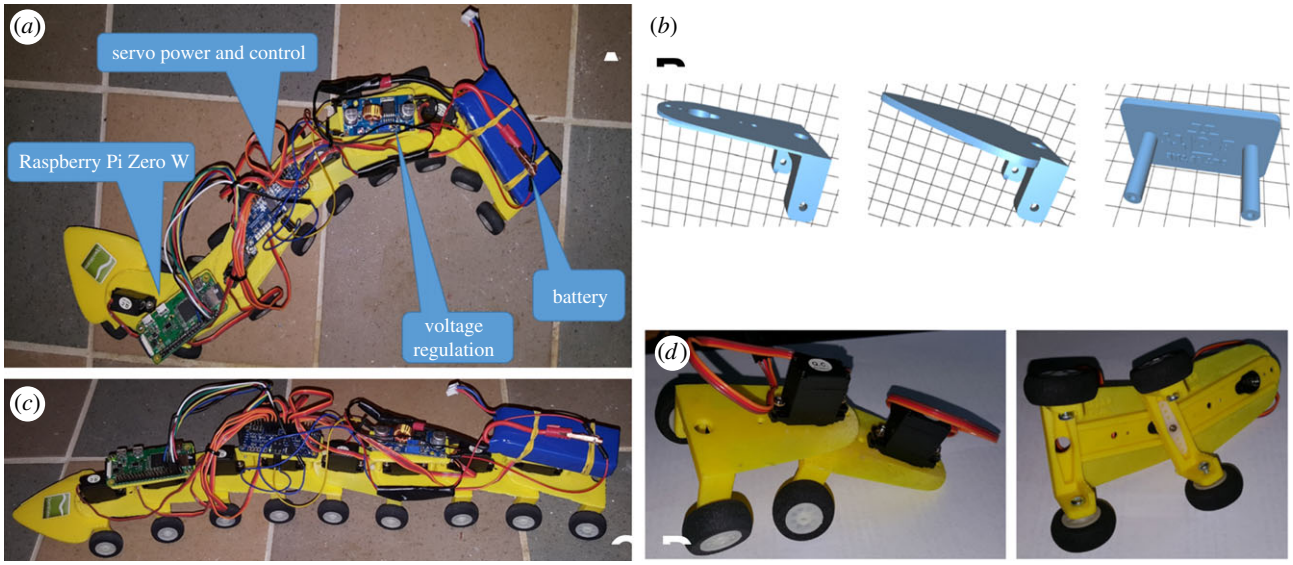

Figure 2. (*a*) Top view of the robot. (*b*) Three-dimensional-printed body parts. (*c*) Side view of the robot. (*d*) Segment sub-assembly.

(iv) Software testing and model validation

As a software project, OpenWorm shares many commonalities with any large-scale software engineering endeavour in industry. *Unit testing* is a key element of modern software engineering which uses semi-automated checklists to ensure the correctness of software. For instance, a company developing a word processor might have a test that verifies whenever the mouse clicks a specific region in the upper left hand of the screen, the 'File' menu opens and not the 'Edit' menu. Likewise, other tests might verify that files can be appropriately written to disk or that connectivity with printers and other network devices is working. From its inception, OpenWorm has incorporated best practices from the software industry, including unit testing, across all of the diverse sub-projects, especially PyOpenWorm [47]. Examples of unit tests used by OpenWorm include verifying that entries can be added to and removed from the PyOpenWorm database, that every biological fact such as ion channel parameters have associated PubMed identifiers and that functions implement error handling correctly.

As a scientific research project that incorporates dynamic models, another class of tests crucial to our effort are *model validation tests*. In contrast to simple unit tests, which verify that a discrete piece of code has the correct behaviour, model validation tests verify that the output of an entire dynamic model corresponds to known behaviour from the academic literature. For instance, alongside the ion channel curation and parameter extraction tasks in ChannelWorm, a parallel effort is aimed at implementing validation tests for each of these models using the Python library SciUnit [48]. The validation process uses curated datasets of ion channel behaviour to instantiate analogous statistical tests that a researcher would use when developing such a model. By incorporating this process into the software development workflow, we can ensure that developers and researchers are alerted if any of the models at any level of abstraction are not in correspondence with known behaviour determined by experimentalists [47,49,50].

(b) Outreach, education and sister projects

(i) Web-based visualization of OpenWorm models

We recognize that many motivated and talented citizen scientists are not experienced in software engineering and data science. Consequently, to make the OpenWorm model as accessible as possible, we have worked to create simple and intuitive applications that can be used for exploratory purposes and which can serve as a fun and compelling entry point to the project. Initial work to accomplish this was the development of the WormSim (http://wormsim.org) prototype. WormSim was launched via a successful Kickstarter campaign in 2014, but this has been superseded by more advanced approaches to visualizing these models. Recent developments with the Geppetto platform (http://geppetto.org) [20] for multi-scale biological simulation, which was the underlying platform for WormSim, have enabled users to visualize the *C. elegans* connectome within the body of the worm itself, and visualize and explore changing dynamics in the connectome to see the effect on swimming and crawling. This version of the visualization is currently being incorporated into the OpenWorm simulation stack above in order to allow users to examine intermediate levels of the simulation (see [20], this issue, for visual examples.)

(ii) Robotics

Because the scientific vision of OpenWorm places a key emphasis on biomechanics, we have multiple outlets for how the virtual nervous system simulation interfaces with the world. One is through a fully virtual body embedded in a virtual physical environment. Another is for the nervous system simulation to interact directly with a robotic body, a platform that provides a unique educational opportunity for newcomers to engage with the project.

Figure 2*a*,*c* shows a top and side view of a prototype OpenWorm robot (https://github.com/openworm/robots) with major components denoted. The robot consists of nine articulated segments, each segment mounted on a pair of wheels. Locomotion is achieved, as it is in *C. elegans*, by moving in a snake-like manner that relies on surface friction. The wheels are not powered and exist solely to provide a suitable contact surface with the ground. Each segment is a three-dimensional-printed component (figure 2*b*) that articulates with its neighbours via servos (figure 2*d*). The electronic components, consisting of the Raspberry Pi Zero microprocessor with wireless communication capabilities, are mounted on platforms fastened to several of the segments. A pulse-width modulation board distributes power and controls signals from the Raspberry Pi Zero to the servos. Each servo is capable of maintaining a specified angular position that translates to inter-segment angular positioning.

Figure 2*b* shows the designs for the 3 three-dimensional-printed parts. These parts are specified in a common .stl file format that is editable and portable to most three-dimensional

printers. On the left is the segment part. In the centre is the head that is envisaged to be mounted with sensors for food foraging and touch. On the right is one of the platforms for mounting the electronic components. Figure 2*d* shows how the segments are articulated. A servo is mounted on the front top of the segment with its geared shaft extending into an aperture in the next forward segment. An arm secured to the gear allows for gear motion to drive angular movement between segments.

Like WormSim, the robotics sub-project of OpenWorm is a key element of our education and outreach efforts. The accessibility and low cost of electronics microprocessors like the Raspberry Pi make this an attractive and compelling introduction to the project for students of all ages, which exposes them to bleeding edge concepts at the intersection of software and robotics. Ongoing work in the robotics sub-project is aimed at incorporating models for food foraging and touch response, developing a new system-on-a-board processor that will also perform power and control distribution, utilizing laser-cut segments and providing a programming interface via Jupyter notebooks.

(iii) DevoWorm

Much of what we have described above pertains to simulations of the adult nematode. A complementary goal for computational research, with direct relevance for many members of the *C. elegans* experimental community, is to simulate embryogenesis and development in *C. elegans*. Given the knowledge of the embryonic cell lineage in *C. elegans* [51], one of the goals of DevoWorm, an ongoing sister project to OpenWorm with a parallel set of approaches (https://github.com/devoworm), is to apply a similar modelling technique of transforming datasets into computable forms and evolving their progression over time using mathematical models of biophysical developmental processes. It is currently divided into three loosely knit sub-projects: Developmental Dynamics, Cybernetics and Digital Morphogenesis, and Reproduction and Developmental Plasticity.

Developmental Dynamics currently involves using secondary data collected from embryos [52,53] along with bioinformatic and data science techniques to answer questions regarding the process of early embryogenesis and the timing of later morphogenesis. Cybernetics and Digital Morphogenesis has involved using cellular automata [54] or finite-element approaches [55] to model physical interactions during embryogenesis and morphogenesis. DevoWorm has also explored the use of cybernetic models and concepts to better understand the general process of embryogenesis [56]. Reproduction and Developmental Plasticity involves an evolutionary developmental biology approach [57] to understand *C. elegans* more generally. DevoWorm's existing datasets and papers include a focus on larval development and life-history processes. Taken together, these focus areas are beginning to draw additional interest into simulated embryogenesis and morphogenesis of *C. elegans*.

(c) Community management support

(i) Distributed scientific collaboration

A citizen science consortium with over 90 contributors[1] from 16 different countries and no central source of funding, OpenWorm has been an organizational experiment in coordinating a distributed, international research effort with a highly fluid base of contributors. Freely available software tools have played a key role in project management and coordination. The focal point of much of our work is the diverse functionality of the GitHub platform [58], which allows us to use sophisticated, industrial-scale management tools for versioning the OpenWorm codebase as well as data from our university-based research partners.

Other platforms such as the Google Docs platform with spreadsheets, drawings, slides and forms have also been critical for the creation and distribution of shared materials. Teleconferencing systems like Google Hangouts have enabled building trust, camaraderie and working relationships among contributors living in many time zones across the globe. Google Calendar has been invaluable for scheduling, as has the Doodle poll tool for coordinating meeting times. The functionality of the Slack chat platform has played a crucial role in managing the many asynchronous conversations related to software development and the scientific roadmap. Given the volume of high-quality tools such as Amazon Web Services, Docker, Slack and many others that are available for use in the modern era of software engineering, the challenges we faced in the initial stages of building the organization often amounted to making the right choices about which tools to use on the basis of the cohesiveness of their relationships with one another. Consequently, the integration points between these different systems have been one of the concrete deliverables of OpenWorm for other organizations interested in distributed community management.

Because of its open source and volunteer-based nature, timelines for task completion are often fluid. Coordinating the project requires the discovery of synergy among collaborators based on individual interests and research goals. Managing the project requires creating the potential for others to contribute and build, connecting that potential to the right individuals at the right time and ensuring that there is sufficient flexibility in the high-level vision so that the project can make progress even if all directions are not advancing at a given moment.

The 'long memory' of online resources is helpful in this regard. Issues that are captured in GitHub may sit inactive for months before the right person comes along who has the skill set and motivation to solve them. Consequently, the tolerance of contributors to uncertainty is an important component of working well within an open community. We have taken inspiration from the open source programming movement that follows a similar philosophy. In open source software development, new volunteers are encouraged to take personal responsibility and leadership for creating new directions that excite them. A unique aspect of research and development in an open community is the rate at which volunteers enter the project eager to learn and to contribute their time and energy to a shared effort that is larger than any one individual [59].

(ii) Mentorship and training through badges

Open-science projects face a very different set of management challenges when compared with university or industry-based research initiatives. In particular, mechanisms are needed to assist new contributors to develop relevant technical skills and build familiarity with the project. To facilitate this process, OpenWorm has taken advantage of a free service called BadgeList, which allows for the creation of digital 'micro-credentials' certifying that an individual is able to complete a focused set of tasks (http://badgelist.com/openworm). Upon successfully learning and answering a set of test questions, a user can earn a badge, indicating that they have acquired a specific skill set. Example badges currently used by the project include basic and advanced GitHub/version control, Hodgkin–Huxley equation basics and literature mining. The collection of badges has been growing over the past several years, and many new contributors have found the system to be a valuable entry point to the project.

(iii) Volunteer composition and project leadership

We have been fortunate that OpenWorm has attracted an incredibly diverse set of volunteers with respect to nationality and intellectual background. As we mentioned above, we have over 90 volunteers from 16 different countries who have made substantive contributions to the project. Moreover, the contributors have come from a variety of academic backgrounds, including theoretical and experimental biology, physics and computer

science, to name just a few. In addition to several core members who are tenure-track faculty at major universities, many of our volunteers are professional software engineers. One area where we are keen to make more progress in is the gender diversity within the project. We have recognized this as a priority and have advertised on social media our active commitment to providing a safe and welcome space for all individuals. We welcome any input on how we might go about achieving a more equitable gender balance.

We are frequently asked about project leadership, decision-making and conflict resolution. Like many open source projects, our list of contributors has a long tail, with a few core contributors assuming leadership roles and many others making periodic, smaller contributions [60,61]. To date, we have had no formal process for assigning roles, and we have found that experienced and enthusiastic volunteers often establish themselves as leaders without any prompting. Subgroups dedicated to topics ranging from engineering, to basic science, to community outreach organize via dedicated channels on Slack, and new volunteers have the opportunity to contribute to whichever efforts resonate with them the most. We have actively worked to ensure a culture where open deliberation takes place with all contributors receiving a voice. With the formal incorporation of the OpenWorm Foundation as an independent, non-profit research organization, we have formed an official scientific advisory board that is responsible for establishing the scientific direction of the effort. Thus far, we have found that input from the scientific advisory board has organically filtered into the project in an effective manner. As the project grows, we may consider formalizing the roles of full-time staff and primary collaborators.

3. Recent progress

What progress has the OpenWorm project made since the publication of our first overview paper [1]? The number of contributors has grown substantially and the codebase has sufficiently matured that new volunteers can join and begin to contribute by tackling open issues on GitHub. Building on several years of experience managing an open-science project, as well as our collective experience building software in the commercial and academic setting, we have refined many of our management practices to better serve the needs of a fluidly shifting base of contributors. The badge system described above has been used by several dozen new members, and we have been holding weekly 'office hours' on Slack where senior contributors are available to answer questions.

With regard to more traditional academic metrics, the special issue in which this article appears will include the publication of several new articles featuring foundational modelling, simulation, data management and data presentation technologies developed as a result of OpenWorm-led collaborations [19,20,23,48]. Before this special issue, we have published a handful of papers on several different facets of the project in a spectrum of journals focused on the computational biological sciences. We have been involved with multiple academic conferences and have built university-based collaborations with six different research laboratories in four countries. Equally as important, we have formally been incorporated as an independent non-profit research organization, the OpenWorm Foundation. This foundation has allowed us to assemble an accomplished scientific advisory board that is helping to guide us through this critical infrastructure-building phase of the project.

To address a query frequently asked of the project: 'when can we turn the simulation on?', the simple answer is that there is already prototype code to do this, available online at GitHub (http://github.com/openworm/openworm). The *C. elegans* connectome contained in c302 is able to drive body movement in the Sibernetic platform for fluid dynamic simulations. However, the level of detail that we have incorporated to date is inadequate for biological research. A key remaining component is to complete the curation and parameter extraction of Hodgkin–Huxley models for ion channels to produce realistic dynamics in neurons and muscles. Once this task is complete, we expect that the platform will incorporate a sufficiently granular level of detail to be of interest to researchers in the field.

Table 2 summarizes our accomplishments.

4. Discussion

By organizing the research output of an entire community into a shared structure, integrative simulations have the potential to advance biological thinking significantly. Rather than being replacements for existing theoretical or experimental techniques, these composite simulations should be viewed as powerful tools to augment the thought process and technical toolbox of scientists. Figure 3 summarizes how integrative simulations can be an organic part of the research process. The same observations that researchers use to form mental models and hypotheses are first organized into databases such as PyOpenWorm and ChannelWorm (arrow *a*). Researchers benefit from these databases directly, for example, by having on-demand access to useful facts about *C. elegans* physiology (arrows *b* and *c*). Subsequently, these datasets are formalized into mathematical models, a process that is itself intrinsically valuable to researchers as part of hypothesis generation (arrow *e*). Most significant are the final steps of this sequence, in which the datasets and mathematical models are integrated into a larger, composite simulation. By studying the outputs of simulations, analogous to the outputs of experiments, researchers are able to augment their intuition and mental models about biological function in ways that would not be possible through experimentation alone (arrows *f* and *g*).

While these agendas are in their nascent stages, many of the key components of the OpenWorm simulation platform will only need to be built once and can then be re-used community-wide in day-to-day research. At the OpenWorm Foundation, we are assembling the necessary technical and organizational infrastructure to build the world's first integrative biological simulation of the nematode *C. elegans*. We hope that subsequent efforts will benefit from our experience, and, in the future, we hope to see the vision of integrative biological simulations extend to many other model organisms and have a widespread scientific impact.

(a) Future directions

Looking forward, there are two thrusts for the project: a primarily scientific one, and the second, a primarily engineering or tool-building phase.

The tap withdrawal circuit is a well-studied experimental protocol we are currently investigating that has focused our transition from infrastructure development to actively using the platform for scientific research [62,63]. Simulating this behaviour will require closing the loop between sensation, motor output and environmental activity. In addition, the

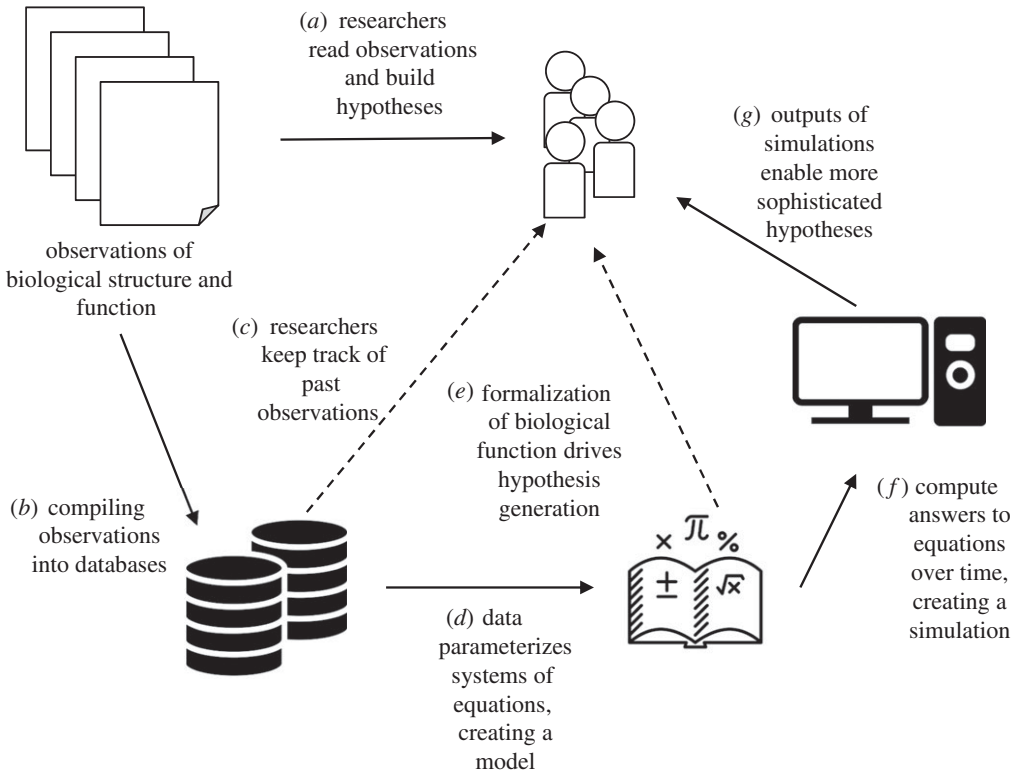

Figure 3. Schematic view of the scientific value of modelling and simulation in biology, as applied in OpenWorm.

Table 2. Recent achievements of the project.

result type	accomplishments
scientific communication	'Connectome to Behavior' conference at The Royal Society, London, UK, 2018
	Workshop at Neural Information Processing Systems (NIPS), Los Angeles, USA, 2017
	Genetics Society of America 22nd International *C. elegans* Conference Workshop, Los Angeles, USA, 2017
community efforts	Office hours—a weekly meeting on Slack open to anyone where senior contributors are available to answer questions about the project
	Badges—16 badges providing 'micro-credentials' for key skills necessary for contributing to OpenWorm. Fifty-one badges have been earned by contributors since the beginning of the project.
	Journal clubs—YouTube-based series reviewing scientific papers relevant to modelling *C. elegans*
	Mailing list—1600 subscribed members
distributed project management tools	8 612 788 lines of code in 51 different programming languages spanning 63 sub-projects (repositories) in GitHub
	Slack workspace has 171 members, 43 weekly active users across 27 public channels
	Twitter account has 3000 followers, average monthly impressions: 25 000, maximum impressions: 46 000 in January 2018.
organizational	OpenWorm Foundation incorporated as a 501c(3) in the USA in 2016. Formed a formal board of directors and a scientific advisory board (http://openworm.org/people.html).
academic collaborations	University College London, London, UK
	Imperial College, London, UK
	Arizona State University, AZ, USA
	A.P. Ershov Institute of Informatics Systems, Novosibirsk, Russia
	TU Wien, Vienna, Austria
	Emory University School of Medicine, GA, USA
publications	Four new publications from OW contributors and collaborators in submission for special issue at *Phil. Trans. Royal Soc.*
	Previous articles listed at: http://openworm.org/publications.html

nervous system model must be able to transform an external input into a switch of behaviour from crawling forwards to backwards. As a prerequisite, we must also implement a version of forward and backward locomotion based on the activity of motor neurons driving the muscles of the model. To ensure the correctness of such a model, we are

incorporating SciUnit-based model validation tests, which allow us to constrain the simulation to match experimental data at different scales and modalities. A critical component of this research direction will be efficient optimization algorithms to help fill in data gaps of free parameters that are currently unknown within the biological community. Once a working prototype of tap withdrawal is completed, we can look at perturbing the model in ways that are consistent with mutations known to have an impact on neuronal or other cellular activity. This will be a valuable test of the ability of the OpenWorm integrated model to capture essential dynamics despite significant biological variation.

Our engineering aim at present is to reach a steady state where the fundamental infrastructure of OpenWorm has stabilized and can be used for scientific research. Active ongoing infrastructural development in the project includes expanding the functionality of PyOpenWorm to store metadata and provenance of simulations, using this framework to build a database of simulation results, completing the ion channel curation and parameter extraction tasks in ChannelWorm, building an automated system for identifying new publications on *C. elegans* relevant for OpenWorm and expanding the automated framework for verifying the correctness of curated scientific models, to name just a few. More information about making a contribution is available on our website and via our volunteer contribution form (http://bit.ly/OpenWormVolunteer).

Data accessibility. All data and code associated with OpenWorm is available through our GitHub repository at https://github.com/OpenWorm.

Authors' contributions. G.P.S. edited and contributed a significant amount of original text. S.D.L. wrote the first draft and provided editing guidance. C.W.L., P.G., D.L., R.M.H. and R.C.G. provided important comments to edit and shape the paper. S.G. designed and built the initial robot, including the three-dimensional-printed parts. T.P. reproduced the design and provided software for running the simulation that produces run data for the robot. T.P., V.G. and T.J. wrote the initial draft of individual sections that they have worked on. Additionally, all the authors have provided edits and have contributed source code to repositories that are noted throughout the document, available at http://github.com/openworm/.

Competing interests. S.D.L., M.C. and G.I. are also co-founders of MetaCell LLC, LTD, a software company that has made financial and in-kind donations to the OpenWorm project pre-2016 and the OpenWorm Foundation since its inception.

Funding. We are grateful to the organizations that have made OpenWorm possible through their financial and in-kind contributions, including Amazon Web Services, Google Summer of Code, NeuroLinx, Open Source Brain (through Wellcome Trust grant 101445), National Institutes of Health (R01MH106674) and National Institute of Biomedical Imaging and Bioengineering (R01EB021711).

Acknowledgements. G.P.S. thanks Victor Faundez and Maureen Powers at the Emory University School of Medicine for encouraging this research. Most of all, none of this would be possible without the many dedicated and energetic OpenWorm Contributors listed at http://openworm.org/people.html.

Endnote

[1]Contributors are defined as any individual who has updated code in our open source repositories or contributed organizationally or scientifically. See http://openworm.org/people.html.

References

1. Szigeti B *et al.* 2014 OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*. *Front Comput. Neurosci.* **8**, 137. (doi:10.3389/fncom.2014.00137)
2. Gleeson P, Cantarelli M, Currie M, Hokanson J, Idili G, Khayrulin S, Palyanov A, Szigeti B, Larson S. 2015 The OpenWorm Project: currently available resources and future plans. *BMC Neurosci.* **16**, P141. (doi:10.1186/1471-2202-16-S1-P141)
3. Harel D. 2004 A grand challenge for computing: towards full reactive modeling of a multi-cellular animal. In *Verification, model checking, and abstract interpretation* (eds B Steffen, G Levi), pp. 323–324. Berlin, Germany: Springer
4. Avgoustinov N. 2007 *Modelling in mechanical engineering and mechatronics: towards autonomous intelligent software models*. London: Springer. See https://market.android.com/details?id=book-dLVrtwAACAAJ.
5. Sarma GP, Faundez V. 2017 Integrative biological simulation praxis: considerations from physics, philosophy, and data/model curation practices. *Cell Logist.* **7**, e1392400. (doi:10.1080/21592799.2017.1392400)
6. Karr JR, Sanghvi JC, Jacobs JM, Macklin DN, Covert MW. 2012 A whole cell model of mycoplasma genitalium elucidates mechanisms of bacterial replication. *Biophys. J.* **102**, 731a. (doi:10.1016/j.bpj.2011.11.3967)
7. Gjorgjieva J, Biron D, Haspel G. 2014 Neurobiology of *Caenorhabditis elegans* locomotion: where do we stand? *Bioscience* **64**, 476–486. (doi:10.1093/biosci/biu058)
8. White JG, Southgate E, Thomson JN, Brenner S. 1986 The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340. (doi:10.1098/rstb.1986.0056)
9. Pereira L *et al.* 2015 A cellular and regulatory map of the cholinergic nervous system of *C. elegans*. *eLife* **4**, 299. (doi:10.7554/eLife.12432)
10. Gendrel M, Atlas EG, Hobert O. 2016 A cellular and regulatory map of the GABAergic nervous system of *C. elegans*. *eLife* **5**, 1395. (doi:10.7554/eLife.17686)
11. Bentley B, Branicky R, Barnes CL, Chew YL, Yemini E, Bullmore ET, Vértes PE, Schafer WR. 2016 The multilayer connectome of *Caenorhabditis elegans*. *PLoS Comput. Biol.* **12**, e1005283. (doi:10.1371/journal.pcbi.1005283)
12. Le Novere N *et al.* 2006 BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* **34**, D689–D691. (doi:10.1093/nar/gkj092)
13. Ascoli GA, Donohue DE, Halavi M. 2007 NeuroMorpho.Org: a central resource for neuronal morphologies. *J. Neurosci.* **27**, 9247–9251. (doi:10.1523/JNEUROSCI.2055-07.2007)
14. Milyaev N, Osumi-Sutherland D, Reeve S, Burton N, Baldock RA, Armstrong JD. 2012 The Virtual Fly Brain browser and query interface. *Bioinformatics* **28**, 411–415. (doi:10.1093/bioinformatics/btr677)
15. Tytell ED, Holmes P, Cohen AH. 2011 Spikes alone do not behavior make: why neuroscience needs biomechanics. *Curr. Opin. Neurobiol.* **21**, 816–822. (doi:10.1016/j.conb.2011.05.017)
16. Hay N, Stark M, Schlegel A, Wendelken C, Park D, Purdy E, Silver T, Phoenix DS, George D. 2018 *Behavior is everything—towards representing concepts with sensorimotor contingencies*. In *32nd AAAI Conf. Artif. Intell., New Orleans, LA, 2–7 February 2018*, pp. 1861–1870. See https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16413. Palo Alto, CA: AAAI Press.
17. Palyanov A, Khayrulin S, Larson SD. 2016 Application of smoothed particle hydrodynamics to modeling mechanisms of biological tissue. *Adv. Eng. Softw.* **98**, 1–11. (doi:10.1016/j.advengsoft.2016.03.002)
18. Palyanov A, Khayrulin S, Larson SD. 2018 Three-dimensional simulation of the *Caenorhabditis elegans* body and muscle cells in liquid and gel environments for behavioural analysis. *Phil. Trans. R. Soc. B* **373**, 20170376. (doi:10.1098/rstb.2017.0376)
19. Gleeson P, Lung D, Grosu R, Hasani R, Larson SD. 2018 c302: a multi-scale framework for modelling the nervous system of *Caenorhabditis*

elegans. *Phil. Trans. R. Soc. B* **373**, 20170379. (doi:10.1098/rstb.2017.0379)

20. Cantarelli M, Marin B, Quintana A, Earnshaw M, Court R, Gleeson P, Dura-Bernal S, Silver RA, Idili G. 2018 Geppetto: a reusable modular open platform for exploring neuroscience data and models. *Phil. Trans. R. Soc. B* **373**, 20170380. (doi:10.1098/rstb.2017.0380)
21. Masoli S, Rizza MF, Sgritta M, Van Geit W, Schürmann F, D'Angelo E. 2017 Single neuron optimization as a basis for accurate biophysical modeling: the case of cerebellar granule cells. *Front. Cell Neurosci.* **11**, 71. (doi:10.3389/fncel.2017.00071)
22. Gurkiewicz M, Korngreen A. 2005 A numerical approach to ion channel modelling using whole-cell voltage-clamp recordings and a genetic algorithm. *PLoS Comput. Biol.* **3**, e169. (doi:10.1371/journal.pcbi.0030169.eor)
23. Javer A, Ripoll-Sánchez L, Brown AEX. 2018 Powerful and interpretable behavioural features for quantitative phenotyping of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. B* **373**, 20170375. (doi:10.1098/rstb.2017.0375)
24. de Bono B, Hunter P. 2012 Integrating knowledge representation and quantitative modelling in physiology. *Biotechnol. J.* **7**, 958–972. (doi:10.1002/biot.201100304)
25. Harris TW, Antoshechkin I, Bieri T, Blasiar D. 2009 WormBase: a comprehensive resource for nematode research. *Nucleic Acids* **38**, D463–D467. (doi:10.1093/nar/gkp952)
26. Altun ZF, Hall DH. 2002 WormAtlas. See http://www wormatlas org.
27. Altun ZF, Herndon LA, Wolkow CA, Crocker C, Lints R, Hall DH (eds). 2002–2018 WormAtlas. See http://www.wormatlas.org.
28. Bargmann CI. 1998 Neurobiology of the *Caenorhabditis elegans* genome. *Science* **282**, 2028–2033. (doi:10.1126/science.282.5396.2028)
29. Goodman MB, Ernstrom GG, Chelur DS, O'Hagan R, Yao CA, Chalfie M. 2002 MEC-2 regulates *C. elegans* DEG/ENaC channels needed for mechanosensation. *Nature* **415**, 1039–1042. (doi:10.1038/4151039a)
30. Strange K. 2003 From genes to integrative physiology: ion channel and transporter biology in *Caenorhabditis elegans*. *Physiol. Rev.* **83**, 377–415. (doi:10.1152/physrev.00025.2002)
31. Salkoff L, Wei AD, Baban B, Butler A, Fawcett G, Ferreira G, Santi CM. 2005 Potassium channels in *C. elegans*. *WormBook* 1–15. (doi:10.1895/wormbook.1.42.1)
32. Hobert O. 2005 Specification of the nervous system. *WormBook* 1–19.
33. Bianchi L, Driscoll M. 2006 Heterologous expression of *C. elegans* ion channels in *Xenopus* oocytes. *WormBook* 1–16.
34. Liu P, Chen B, Wang Z-W. 2014 SLO-2 potassium channel is an important regulator of neurotransmitter release in *Caenorhabditis elegans*. *Nat. Commun.* **5**, 5155. (doi:10.1038/ncomms6155)
35. Hodgkin AL, Huxley AF. 1952 A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544. (doi:10.1113/jphysiol.1952.sp004764)
36. Willms AR, Baro DJ, Harris-Warrick RM, Guckenheimer J. 1999 An improved parameter estimation method for Hodgkin-Huxley models. *J. Comput. Neurosci.* **6**, 145–168. (doi:10.1023/A:1008880518515)
37. Boyle JH, Cohen N. 2008 *Caenorhabditis elegans* body wall muscles are simple actuators. *Biosystems* **94**, 170–181. (doi:10.1016/j.biosystems.2008.05.025)
38. O'Leary T, Williams AH, Franci A, Marder E. 2014 Cell types, network homeostasis, and pathological compensation from a biologically plausible ion channel expression model. *Neuron* **82**, 809–821. (doi:10.1016/j.neuron.2014.04.002)
39. Mirzakhalili E, Epureanu B, Gourgou E. 2017 A mathematical and computational model of the calcium dynamics in *Caenorhabditis elegans* ASH sensory neuron. *PLoS ONE* **13**, e0201302. (doi:10.1371/journal.pone.0201302)
40. Kuramochi M, Doi M. 2017 A computational model based on multi-regional calcium imaging represents the spatio-temporal dynamics in a *Caenorhabditis elegans* sensory neuron. *PLoS ONE* **12**, e0168415. (doi:10.1371/journal.pone.0168415)
41. Markram H *et al.* 2015 Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492. (doi:10.1016/j.cell.2015.09.029)
42. Ranjan R, Khazen G, Gambazzi L, Ramaswamy S, Hill SL, Schürmann F, Markram H. 2011 Channelpedia: an integrative and interactive database for ion channels. *Front. Neuroinform.* **5**, 36. (doi:10.3389/fninf.2011.00036)
43. Milescu LS, Akk G, Sachs F. 2005 Maximum likelihood estimation of ion channel kinetics from macroscopic currents. *Biophys. J.* **88**, 2494–2515. (doi:10.1529/biophysj.104.053256)
44. Wang W, Xiao F, Zeng X, Yao J, Yuchi M, Ding J. 2012 Optimal estimation of ion-channel kinetics from macroscopic currents. *PLoS ONE* **7**, e35208. (doi:10.1371/journal.pone.0035208)
45. Gleeson P *et al.* 2010 NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail. *PLoS Comput. Biol.* **6**, e1000815. (doi:10.1371/journal.pcbi.1000815)
46. Cannon RC, Gleeson P, Crook S, Ganapathy G, Marin B, Piasini E, Silver RA. 2014 LEMS: a language for expressing complex biological models in concise and hierarchical form and its use in underpinning NeuroML 2. *Front. Neuroinform.* **8**, 79. (doi:10.3389/fninf.2014.00079)
47. Sarma GP, Jacobs TW, Watts MD, Ghayoomie SV, Larson SD, Gerkin RC. 2016 Unit testing, model validation, and biological simulation. *F1000Res.* **5**, 1946. (doi:10.12688/f1000research.9315.1)
48. Gerkin RC, Jarvis RJ, Crook SM. 2018 Towards systematic, data-driven validation of a collaborative, multi-scale model of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. B* **373**, 20170381. (doi:10.1098/rstb.2017.0381)
49. Gerkin RC, Omar C. 2013 NeuroUnit: validation tests for neuroscience models. *Front. Neuroinform.* (doi:10.3389/conf.fninf.2013.09.00013)
50. Omar C, Aldrich J, Gerkin RC. 2014 *Collaborative infrastructure for test-driven scientific model validation*. In *Companion Proc. of the 36th Int. Conf. Software Engineering*, pp. 524–527. ACM.
51. Sulston JE. 1983 Neuronal cell lineages in the nematode *Caenorhabditis elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **48**, 443–452. (doi:10.1101/SQB.1983.048.01.049)
52. Santella A *et al.* 2015 WormGUIDES: an interactive single cell developmental atlas and tool for collaborative multidimensional data exploration. *BMC Bioinformatics* **16**, 189. (doi:10.1186/s12859-015-0627-8)
53. Wang E, Santella A, Wang Z, Wang D, Bao Z. 2017 Visualization of 3-dimensional vectors in a dynamic embryonic system—WormGUIDES. *J. Comput. Commun.* **5**, 70–79. (doi:10.4236/jcc.2017.512008)
54. Portegys T, Pascualy G, Gordon R, McGrew SP, Alicea BJ. 2017 Morphozoic, cellular automata with nested neighborhoods as a metamorphic representation of morphogenesis. In *Multi-agent-based simulations applied to biological and environmental systems*. (ed. DF Adamatti), pp. 44–80. Hershey, PA, USA: IGI Global.
55. Izaguirre JA *et al.* 2004 CompuCell, a multi-model framework for simulation of morphogenesis. *Bioinformatics* **20**, 1129–1137. (doi:10.1093/bioinformatics/bth050)
56. Gordon R, Stone R. 2017 Cybernetic embryo. In *Biocommunication* (eds R Gordon, J Seckbach), pp. 111–164. London, UK: World Scientific.
57. Carroll SB. 2005 *Endless forms most beautiful: the new science of Evo devo and the making of the animal kingdom*. WW Norton & Company. See https://market.android.com/details?id=book-CnnGKjzw3xMC.
58. Perez-Riverol Y *et al.* 2016. Ten simple rules for taking advantage of Git and GitHub. *PLoS Comput Biol*. **12**: e1004947. (doi:10.1371/journal.pcbi.1004947)
59. Nielsen M. 2012 *Reinventing discovery: the new era of networked science*. Princeton University Press. See https://market.android.com/details?id=book-afqfFW8WV9cC.
60. Lerner J, Tirole J. 2003 Some simple economics of open source. *J. Ind. Econ.* **50**, 197–234. (doi:10.1111/1467-6451.00174)
61. Fang Y, Neufeld D. 2009 Understanding sustained participation in open source software projects. *J. Manage. Inf. Syst.* **25**, 9–50. (doi:10.2753/MIS0742-1222250401)
62. Wicks SR, Rankin CH. 1995 Integration of mechanosensory stimuli in *Caenorhabditis elegans*. *J. Neurosci.* **15**, 2434–2444. (doi:10.1523/JNEUROSCI.15-03-02434.1995)
63. Wicks SR, Roehrig CJ, Rankin CH. 1996 A dynamic network simulation of the nematode tap withdrawal circuit: predictions concerning synaptic function using behavioral criteria. *J. Neurosci.* **16**, 4017–4031. (doi:10.1523/JNEUROSCI.16-12-04017.1996)