JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Research

Cite this article: Hall BA, Jackson E, Hajnal A, Fisher J. 2014 Logic programming to predict cell fate patterns and retrodict genotypes in organogenesis. *J. R. Soc. Interface* **11**: 20140245.
http://dx.doi.org/10.1098/rsif.2014.0245

Received: 7 March 2014
Accepted: 5 June 2014

Subject Areas:
computational biology, systems biology

Keywords:
executable modelling, logic programming, *C. elegans*, organogenesis, development

Authors for correspondence:
Benjamin A. Hall
e-mail: benhall@microsoft.com
Jasmin Fisher
e-mail: jasmin.fisher@microsoft.com

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2014.0245 or via http://rsif.royalsocietypublishing.org.

Royal Society Publishing

Logic programming to predict cell fate patterns and retrodict genotypes in organogenesis

Benjamin A. Hall[1], Ethan Jackson[2], Alex Hajnal[3] and Jasmin Fisher[1,4]

[1]Microsoft Research, 21 Station Road, Cambridge CB1 2FB, UK
[2]Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
[3]Institute of Molecular Life Sciences, University of Zurich, Zurich 8057, Switzerland
[4]Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK

Caenorhabditis elegans vulval development is a paradigm system for understanding cell differentiation in the process of organogenesis. Through temporal and spatial controls, the fate pattern of six cells is determined by the competition of the LET-23 and the Notch signalling pathways. Modelling cell fate determination in vulval development using state-based models, coupled with formal analysis techniques, has been established as a powerful approach in predicting the outcome of combinations of mutations. However, computing the outcomes of complex and highly concurrent models can become prohibitive. Here, we show how logic programs derived from state machines describing the differentiation of *C. elegans* vulval precursor cells can increase the speed of prediction by four orders of magnitude relative to previous approaches. Moreover, this increase in speed allows us to infer, or 'retrodict', compatible genomes from cell fate patterns. We exploit this technique to predict highly variable cell fate patterns resulting from *dig-1* reduced-function mutations and *let-23* mosaics. In addition to the new insights offered, we propose our technique as a platform for aiding the design and analysis of experimental data.

1. Introduction

Formal models of biology integrate data from different experimental sources to rigorously model biological phenomena [1–3]. State-based approaches specifically have been successfully applied to model a wide range of biological systems, including pancreatic organogenesis [4], T-cell differentiation [5,6], vulval development [7–9] and gene regulatory networks [10,11]. One of the advantages of these approaches is their ability to prove a given result. This allows us to find all possible outcomes resulting from a state-based model without attempting an exhaustive search of all executions. Such model analysis, which can normally only be applied to discrete models, is called *symbolic model checking* [12], and allows modellers to identify each possible outcome with certainty. By combining model execution and model checking, it is possible to prove properties of biological hypotheses and compare them with experimental data which subsequently validate or invalidate the hypothesis [13]. These techniques allow us to thoroughly explore experimental datasets and highlight gaps in our understanding of biological mechanisms.

Cell fate determination in *Caenorhabditis elegans* vulval development is a paradigm system for understanding organogenesis in multi-cellular eukaryotes [14,15]. The interplay of two competing signalling systems, the inductive signal cascade (LET-23) and the LIN-12/Notch signalling system, determines the patterning of three cell fates in the six vulval precursor cells (VPCs; figure 1 and box 1). This patterning is invariant in wild-type animals, and can be disrupted by mutations in different parts of the pathways. Previously, a diagrammatic model proposed by Sternberg & Horvitz [8,16] was formalized and refined to describe independent cellular timings and a wide range of different observed mutations [7,9]. A key finding here was the role of bounded asynchrony in

Box 1. Terminology.

substance	lowercase names in italics refer to a gene or genes (e.g. *let-23*, *lst*), whereas a non-italicized name in uppercase refers to an expressed protein (e.g. LET-23)
genotype	the states of a set of genes in an animal
concurrency	the simultaneous and near-simultaneous behaviours of systems such as the VPCs
model checking	the in-depth analysis of a model to guarantee a behaviour of that model
components	proteins or genes which are represented as independent entities with their own state
potential states	the set of states a component can adopt

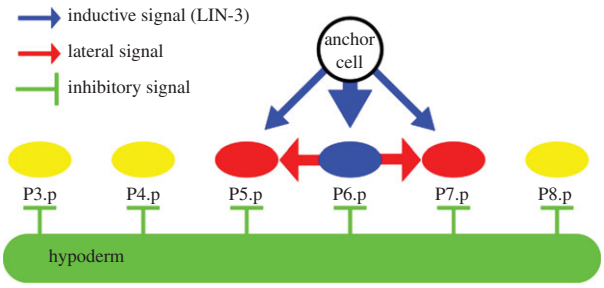

Figure 1. Competition between the inductive and lateral signals determines cell fate. Blue cells are 1°, red 2° and yellow 3°. Strong inductive signal from the anchor cell causes the neighbouring VPC (P6.p) to activate the LET-23/LET-60/MAPK cascade, which leads to the cell generating the lateral signal and the cell adopting the 1° fate. Adjacent cells (P5.p and P7.p) receive the lateral signal (red arrow), which inhibits the transduction of the weaker inductive signal and leads to the cell adopting the 2° cell fate. Other cells (P3.p, P4.p and P8.p) which do not receive any signals adopt the 3° fate. LIN-15 prevents secretion of LIN-3 by the hypoderm (green inhibitory arrow), and a loss of function mutation of LIN-15 causes the cells at P3.p, P4.p and P8.p to assume a non-3° fate.

the development of cell fates [9,17], leading to multiple possible cell fate patterns when the gene *lin-15* is mutated. Bounded asynchrony allows cells to progress semi-independently; this allows individual cells to develop at differing rates, but ensures that the degree of variation is limited by forcing fast developing cells to wait for their neighbours.

Molecular cross-talk between vulval cell fate determination pathways has become increasingly well defined by experimental findings (figure 2). The current view describes the inductive signal cascade beginning with LIN-3 secretion from the anchor cell, to the LET-23/SEM-5/LET-60/MAPK pathway specifying the primary (1°) cell fate. Alongside this, the lateral signal (encoded by three members of the Delta/Serrate protein family [18]) is generated by VPCs in response to MAPK activation, and activates LIN-12/Notch to specify the secondary (2°) fate. The VPC P6.p, located immediately adjacent to the anchor cell, receives the strongest inductive signal and so assumes the 1° fate in the wild-type. Adjacent P5.p and P7.p cells receive a strong lateral signal from P6.p, which inhibits transduction of the inductive signal via activation of several lateral signal target genes (collectively termed *lst* genes) and instead specifies the 2° cell fate. The absence of inductive or lateral signals experienced by other cells leads to their adopting the tertiary (3°) cell fate and the ultimate 3° 3° 2° 1° 2° 3° pattern.

Specific mutations, such as *lin-15* null-alleles (which cause an increase in LIN-3 expression by the surrounding hypodermis), lead to variable outcomes which reflect alternative ordering of molecular and cellular events [16,19]. Neighbouring pairs of 1° cells are rarely observed in *lin-15* mutants, despite all VPCs experiencing high levels of inductive signal [16]. This is explained in terms of individual cells' LET-23/SEM-5/LET-60/MAPK pathway being activated at different times. Early induction of one cell (giving a 1° fate) leads to its neighbours adopting a 2° fate as a result of the 1° cell generating a lateral signal [16]. In a formal model [7], this can be viewed as a result of the specific ordering of independently occurring signalling events. An alternative mutation which induces variable patterning is the loss-of-function mutation of the extracellular matrix protein DIG-1, which induces a dislocation of the gonad with the anchor cell in mutant worms [20]. The severity of this dislocation differs in individuals carrying the *dig-1* mutation; approximately 90% of animals have a ventrally located gonad, with a variable anterior shift. However, approximately 10% of animals show a dorsal gonad, and a greater variability in cell fate. Surprisingly, dorsal gonads are still able to induce non-3° cell fates, despite placing the anchor cell at least 20 µm from the nearest VPC (in wild-type animals cells greater than 20 µm uniformly achieve the 3° fate). Others have shown that the absolute quantity of LIN-3 is limiting, based on the insensitivity of cell fate patterning to increased copy numbers of LET-23 [21]. Together, these results indicate that patterns observed in *dig-1* mutants arise from a limited quantity of LIN-3 diffusing freely in the body of the worm, causing individual cells to experience variable amounts of inductive signal, regardless of their relative position to the anchor cell. Modelling such a system would require 4096 different combinations of signal strengths to fully explore the resulting consequences (resulting from six cells which may each experience one of four different signal strengths, giving 4^6 possibilities). This would require significant computational resources, owing to the large number of possible inputs and the large numbers of possible interleavings (the model checking of a single mutation requires up to 10 GB RAM and 5 h).

Here, we apply a novel conceptualization to overcome these limitations for the analysis of biological models. We describe the system as a logic program, implemented in the language FORMULA [22–25], and use this to propose cell fate patterns. Each component of the system has a defined, finite set of potential states (the total number of states reflecting the sensitivity of the system), and rules which apply across the system, governing how individual states relate to one another. The locations of each VPC in the row and individual time points are considered as a set of states defined by integer values (line 19 in ElegansDev.4ml), and rules can define relationships between them and proteins or genes in the system. To find a fate pattern from the model, we pose

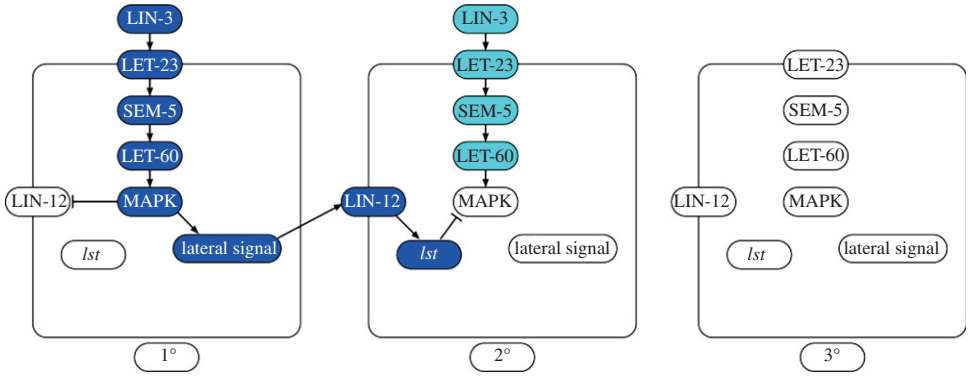

Figure 2. Specific pathways of cell fate determination. Active elements are indicated in dark blue, partially active elements in cyan, and inactive elements in white. Activating relationships are indicated by an arrow, inhibitory by a blocked line. Activation of LET-23 by LIN-3 leads to downstream activation of MAPK, which generates the lateral signal (via LIN-1), and inhibits LIN-12 (via SUR-2). The lateral signal activates LIN-12 in the adjacent cells, which inhibits the LET-23/SEM-5/LET-60/MAPK cascade through multiple lateral signal target gene products (collectively termed *lst* genes). In 3° cells, neither pathway is activated. LIN-15 loss-of-function mutations cause LIN-3 to be secreted by the hypoderm, strongly activating LET-23 in all six VPCs.

a query in the form of a blank table of different possible states in the system over time, where the states of some or all entities are unknown (see www.riseforfun.com/Formula/slmcy for a simple example). A user defines a set of constraints in the logic program to describe the relationships between proteins, genes and time. This information includes both information about the number of states a protein or gene has (typically 2–4) and how the state of the gene and other proteins in the network determine the level of activity. They then define a partially complete table (or query), representing the state of the whole system over a period of time. In a query, each row represents a different element in the model, and each element contains a defined set of information about that element. For example, this can be a cell, containing information on the cell location, the time and the states of all of the proteins within the cell. Alternatively, the element could be a genome, and contain all the gene states used in the model (see the electronic supplementary material, figure S1, for further details). Based on these rules and supplied input information (i.e. the states of different defined genes), we use a model checker, integrated into FORMULA, to complete partial queries (i.e. to offer a single state for the system, or collection of states compatible with the constraints and input). Thus, a user can define a single genotype in a query, and find all possible phenotypes which may arise from that genotype. In the simple example, the resulting 'all facts' table represents a single complete outcome, including the resulting phenotype. Our approach allows us to predict the results of a given mutation more quickly than previous approaches, by up to four orders of magnitude. The model queries are insensitive to the direction of time (i.e. initial states can be used to find final states and vice versa). This allows the model to both predict the phenotype which arises from a given genotype, as well as a genotype that will give rise to a given phenotype. In other fields, this inference of previous events from an observed state has been termed retrodiction [26,27]. Retrodiction in this sense offers a powerful new tool for experimental analysis and biomedical diagnosis, allowing the user to propose compatible genotypes from a given phenotype in terms of known behaviours. Coupled with the increase in speed, this approach could enable high-throughput classification of experimental systems. Furthermore, it offers the potential for specification of cell fate by design, through the proposal of putative mutant combinations for a desired behaviour.

Here, we apply our new approach to modelling *C. elegans* vulval development. The development of the vulva from a row of six cells is a paradigm for understanding organogenesis, and has been modelled previously, using executable [7–9,28], mathematical [29,30] and hybrid models [31]. We initially validate our approach by comparing our models' ability to accurately reproduce the behaviours of previous formal models. We further apply and refine our new models to study the consequences of *dig-1* loss-of-function mutations [20] and LET-23 mosaics, neither of which could have been previously explored owing to the size of the initial state space. We find that while our original model was initially capable of reproducing 75% of fate patterns in *dig-1* loss-of-function mutations, we find that our model requires four refinements to allow all observations to be generated. Specifically, (i) the receiving cell responds to the summation of lateral signals from multiple cells, (ii) lateral signals between cells are maintained after G1 phase, (iii) P3.p and P8.p uniquely are capable of autocrine signalling, and (iv) a low level of MAPK induction results in a weak generation of lateral signal but not a 1° cell fate. These model refinements are sufficient to allow generation of the observed *dig-1* loss-of-function fates. We further apply our technique to *let-23* mosaic animals [21,32,33], where only a subset of the VPCs lack LET-23, and find that modelling of the diffusion of signal from cells with *let-23* loss-of-function mutations is necessary to predict all observed cell fate patterns. Overall, our model demonstrates the requirement of a controlled, polar diffusion of a limited quantity of LIN-3 from the AC, and its subsequent endocytosis on binding to LET-23, to achieve correct cell fate patterns in wild-type animals. Furthermore, to generate the cell fate patterns observed in *dig-1* mutants, there are at least four distinct levels of MAPK activation in VPCs. Finally, our revisions now enable us to reconcile our state-based models with historical data on *lin-3* expression and observations not previously captured [16,34,35].

2. Methods

2.1. Model design and implementation

FORMULA models consist of a specification, written as a logic program and describing all objects in the system in terms of their available states and their relationships, and a query,

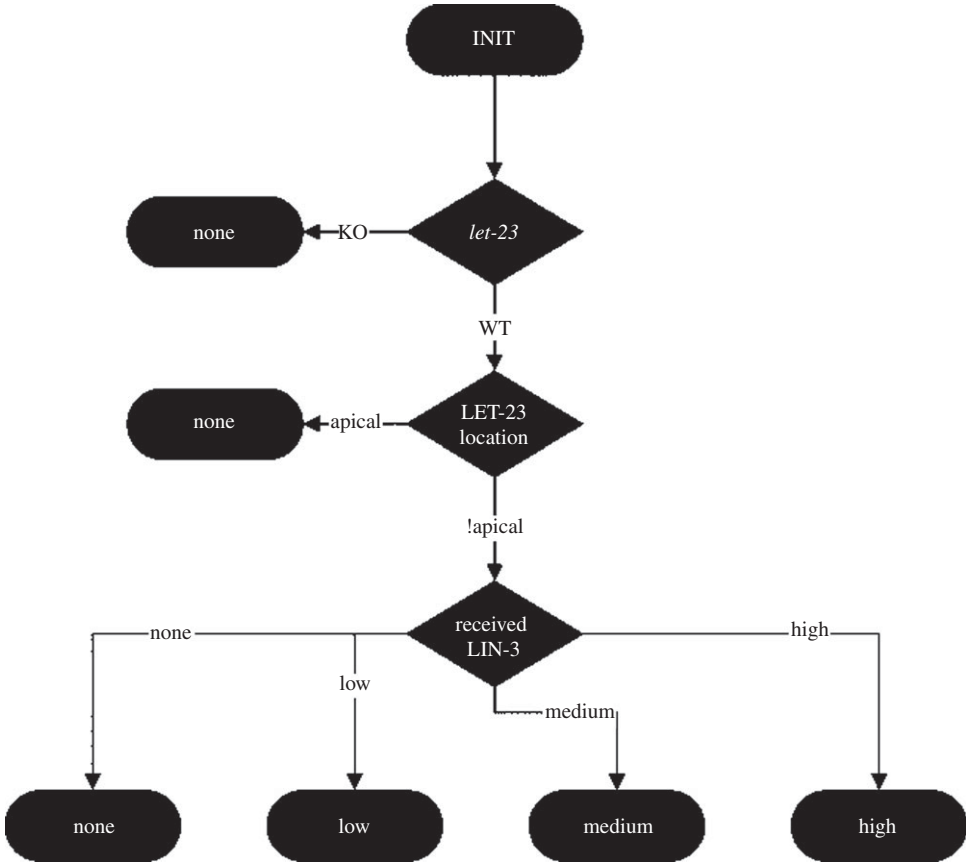

Figure 3. Flowchart shows how LET-23 activity varies. The activity of LET-23 is dependent on a combination of the state of its gene, its location (determined by the state of *lin-2*), and the level of LIN-3 the cell is responding to. If *let-23* is lost, or the location of LET-23 is in the apical membrane, then the activity of LET-23 is always set to none. When LET-23 is located mostly or entirely in the basolateral membrane, its activity matches the level of LIN-3 experienced by the cell; when its location is apical, its activity is always none.

containing a set of known and unknown facts about a specific system. For a given query and specification, FORMULA proposes possible complete models (if they exist), describing all elements in the query explicitly. Specifically, FORMULA achieves this by reducing the logic programming language to query operation for the Z3 constraint solver. In model cells (the anchor cell, VPCs and the hypoderm, figure 1), the genome and 'diffusion locations' (representing areas neighbouring the cells where free LIN-3 may diffuse) are included as objects which can be queried. Each object contains elements which describe individual variables, for example, the state of LET-23 in the VPC, or the mutation of *lin-12* in the genome. Additionally, the VPCs include elements describing time and space, which can be included in other rules, for example, the lateral signal alters the activity of LIN-12 only in neighbouring cells. A typical query therefore contains one genome, one anchor cell, 6×3 VPCs (for each VPC location in the G1, S or G2/M cell cycle phase), one hypodermal multi-nucleate cell and six diffusion locations (electronic supplementary material, figure S1). To simplify query analysis, we can include objects that capture just the final fate patterns of the VPCs, and can explicitly prevent certain fates from occurring (thus proposing alternative fate patterns).

To encode the relationships between elements (such as proteins) within cells, we specify all possible outcomes as constraints. For example, the activity of LET-23 is controlled by the state of its associated gene *let-23* (either wild-type or loss of function), and the cellular level of LIN-3. This allows us to create rules which specify how the level of LET-23 varies in relation to these two elements. For example, if *let-23* (the gene) is lost, LET-23 (the protein) is always inactive (line 227 in EleganSDev.4ml), but if *let-23* is wild-type, rules specify how the activity of LET-23 varies depends on the level of LIN-3 (figure 3, lines 231–236 in EleganSDev.4ml). The level of LIN-3 is determined in turn by the presence or absence of the anchor cell and its location in the worm (i.e. is the gonad dorsal or ventral, and which cell does the anchor cell neighbour, lines 95–98 and 184–202 in EleganSDev.4ml).

Similarly, rules can include notions of time; for example, if a VPC at a given location has experienced a high level of *lst* gene product activity in the past, then this prevents the activity of MAPK from subsequently rising. The more complex behaviour of MAPK is shown as a flow chart in the electronic supplementary material. In G1, the activity of MAPK is dictated by the activity of LET-60 (lines 273–278 in EleganSDev.4ml). After the G1/S checkpoint has been passed, the activity of MAPK is dependent on the events of G1 in addition to the activity of LET-60. If *lst* has never been activated, or if *lst* has previously been successfully inhibited by MAPK activity, then MAPK behaves as it does in G1, and its activity is determined by LET-60 (lines 287–292 in EleganSDev.4ml). If *lst* has been fully activated, and MAPK has not previously successfully inhibited *lst*, then MAPK activity is low if LET-60 is highly active, and non-deterministically will either have medium or no activity if LET-60 is weakly active; for all other levels of LET-60, MAPK is inactive (lines 301–306 in EleganSDev.4ml). If *lst* is partially active (owing to the loss of some of the redundant inhibitors), then the behaviour of MAPK is the same as if *lst* is fully active except when LET-60 is in a sensitized state (owing to loss of function mutations to dep-1) where the activity is non-deterministically selected to be either medium or no activity (lines 294–299 in EleganSDev.4ml).

The rules and granularities of different proteins are expressed as constraints in the logic program by the modeller. Typically, genes have two states (loss-of-function, KO, and wild-type,

WT), though individual genes have additional states to reflect available mutant data (e.g. reduced function for LIN-3 and gain-of-function for LET-60). Proteins have typically two to four states, reflecting the sensitivity required in the models to reproduce the observed fate patterns. There is additionally a set of components that reflect which membrane LET-23 is trafficked to in the cell (either the apical membrane, the basolateral membrane or a mixture). There are 12 genes and 13 proteins in the model. To aid the analysis, fully annotated flow charts representing the behaviour of the specification, alongside a detailed description and specific lines from the code are included in the electronic supplementary material. Queries are insensitive to the information included. This allows the prediction of fate patterns arising from a set of genes, or from a given fate pattern to retrodict compatible sets of genes.

The resulting models as logic programs, compared with models as finite automata (e.g. models written in NuSMV as presented in [9]), can be seen to reproduce the same fate patterns despite the smaller computational costs in model checking. This reduction in cost comes from the loss of events which can be observed in simulations of the finite automata models which do not alter the final outcome of the model. As such, we could potentially lose some mechanistic details regarding the specific ordering of events, if they did not alter the fate pattern. Neither approach is capable at present of proposing probabilities of fate patterns arising in non-deterministic mutants, though it may be possible in future to attach a probability to the non-deterministic events.

Initially, models were generated to reproduce previously observed results [9], before being refined to allow for variable LIN-3 signals in dorsally located *dig-1* mutants, and shifted anchor cells in ventrally located *dig-1* mutants. Multiple redundant inhibitors of the LET-23/SEM-5/LET-60/MAPK pathway (e.g. *sli-1*, *gap-1*) were considered part of the *lst* family for the purpose of the model; a reduced-function form of *lst* was used to model mutations which caused a partial loss of function (e.g. *lip-1*). Experimental observations were then used to validate the new model and highlight inexplicable observations by attempting to construct models with *dig-1* loss-of-function mutations for which the final outcome was known. The concepts of lateral signal permanence and summation, and low level MAPK activation were added to the model, and this process repeated. Finally, rules defining the basis for mosaic mutations were added and fate patterns observed. FORMULA is available at http://research.microsoft.com/en-us/um/redmond/projects/formula/organogenesis2013.html.

The number of possible genotypes (excluding *dig-1* reduced-function and mosaic mutations) totals approximately 10 000, not inclusive of systems where the anchor cell is ablated. This gives rise to up to approximately 630 000 possible executions for the cells. Excluding *lin-15* loss-of-function mutations, which each generate 64 different configurations, this gives approximately 5000 mutant combinations, taking approximately 5 CPU hours to construct. These, shown in the electronic supplementary material, table S1, reproduce behaviour from previous models, and increase the total number of predictions by two orders of magnitude (while reducing calculation time). Large numbers of these mutants share the same cell fates, and several have not yet been characterized experimentally; however, these demonstrate the validity of the approach. Additionally, the speed of the calculations makes larger searches of mutations within a model accessible on a short time scale.

2.2. Concurrency and cellular timings

Individual cells in development are known to adopt specific fates semi-autonomously, and the order in which their fates are adopted can alter the overall fate patterns, giving multiple alternative fate patterns for a given mutation. This arises while the cells are all responding at roughly the same time, as small, stochastic differences in response speeds can lead to different worms adopting different fate patterns, owing to triggering irreversible switches. One mutation that leads to a variable fate pattern is the *lin-15* loss-of-function mutation (lines 135–137, 154–156 in ElegansDev.4ml). This causes the hypodermis to generate high levels of LIN-3 [19,36], making it possible for all VPCs to adopt the 1° fate (though this is only observed rarely [16]). If progression through the process of cell fate determination is modelled as a synchronous, deterministic process, VPCs responding to the hypodermal LIN-3 signal adopt 1° fates exclusively [8]. To allow variable cell fates to arise, cells progress semi-independently (as described by the concept of bounded asynchrony [7,9]). This regime allows individual cells to adopt a 1° fate earlier than their neighbours, and to transmit a lateral signal to their neighbours before they have adopted a 1° fate, causing them to become 2°. More generally, the computational modelling of simultaneous and near-simultaneous executions is termed concurrency. A key feature is that all cells and all mutations share the same mechanism of concurrency, but it is only the *lin-15* mutation which reveals the underlying concurrency by giving a variable fate pattern.

To model the non-deterministic results of *lin-15* mutants, a minimal model of concurrency exploiting this dependency was implemented. Specifically, the model allows each cell to proceed into the 1° state at two different times after the VPCs have responded to inductive signals received from the anchor cell. If at a previous time a cell has not proceeded to the 1° fate, but its neighbour has, it is only able to assume the 2° fate owing to strong lateral signal experienced previously. This allows the generation of multiple different cell fates. In a *lin-15* loss-of-function mutation, this leads to 64 different distinct sets events that can occur, with a smaller number of observed fate patterns (two different timing possibilities for each of six cells, $2^6 = 64$). This non-deterministic behaviour is observed in systems where the outcome in terms of cell fate is uniform (for example, where *lst* is lost as well as *lin-15*), reflecting the fact that each model differs by its individual evolution if not its ultimate outcome. Results of a smaller search of *lin-15* mutant combinations (electronic supplementary material, table S2) reproduce previous observations and behaviours.

3. Results

3.1. *dig-1* mutations cause variable fate patterns through random diffusion of LIN-3

dig-1 encodes a giant (approx. 13 100 residues), multi-domain extracellular matrix protein of the immunoglobulin superfamily [20]. Homozygous loss-of-function mutations of *dig-1* cause major alterations to *C. elegans* physiology by causing defects in the basement membrane structure and disrupt many cell–cell interactions. These effects include neuronal defasicularization [37], nerve axon and cell body displacement [38], and displacement of the gonad [20]. In all worms observed, the gonad was displaced towards the anterior in *dig-1* loss-of-function mutants, and in 10% of worms the gonad was additionally dislocated to the dorsal side of the worm body (figure 4). This dorsal dislocation creates a distance between the anchor cell and its nearest VPC of approximately 20 μm (approximately the distance between 3° cells and the anchor cell in wild-type worms). Despite this distance in worms with dislocated gonads, the anchor cell is able to induce VPCs to adopt 1° or 2° fates at a distance, suggesting that LIN-3 is capable of diffusing across the body. The fate patterns observed in *dig-1* loss-of-function mutants are however highly variable.

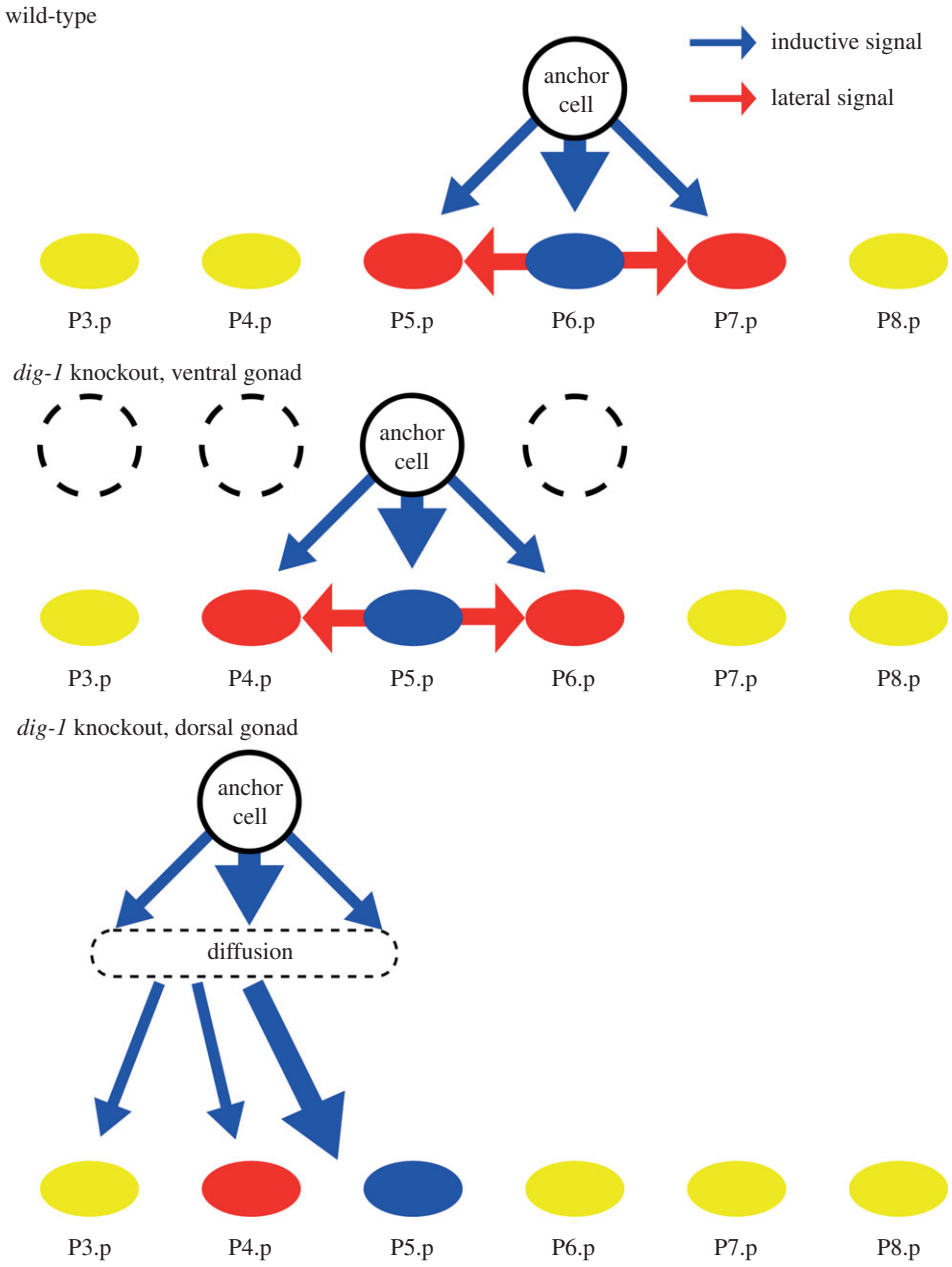

Figure 4. Changes induced by the *dig-1* mutation. In the wild-type, the anchor cell is always ventral and located next to P6.p. In *dig-1* loss-of-function mutants where the gonad is ventral, the anchor cell can be located next to P3.p to P6.p (indicated by dashed cell edges) but otherwise behaves as wild-type. In *dig-1* mutants where the gonad is dorsal, the cells each receive a variable quantity of inductive signal, regardless of anchor cell position.

In worms with ventral gonads, a 2° 1° 2° pattern follows the anterior shift of the anchor cell in the gonad (i.e. if the anchor cell resides above P5.p, the fate pattern becomes 3° 2° 1° 2° 3° 3°). By contrast, worms with dorsal gonads show a greater variety of fate patterns, apparently arising from a random distribution of LIN-3 across the VPCs owing to extracellular diffusion (table 1). Worms with dorsal gonads display wild-type, shifted wild-type (similar to worms with anteriorly displaced, ventrally located gonads) or no defined fate pattern, indicating that the LIN-3 can either diffuse to VPCs generating patterns that mimic wild-type behaviour, or never arrive. In addition, however, more complicated patterns were additionally observed, suggesting that LIN-3 is capable of diffusing to all cells, and that any individual cell could receive any quantity of LIN-3 (within a limit set by the total quantity of LIN-3 secreted).

To model this behaviour, ventral and dorsal variants of *dig-1* loss-of-function mutations were treated differently. *dig-1* loss-of-function mutations with ventral gonads behave as wild-type worms, but the anchor cell is allowed to shift between zero and four cells towards the worm anterior, while maintaining the relationships between LIN-3 delivered to VPCs and its position (i.e. the neighbouring VPC receives a high level of LIN-3, its near neighbour VPC receives a medium level, and distant cells receive a low level). By contrast, VPCs in worms with *dig-1* loss-of-function mutations and dorsally dislocated gonads can receive any level of inductive signal, as long as the sum of the inductive signal received is below a predefined maximum. In our initial models, these new behaviours were sufficient to reproduce all but one cell fate in worms with ventral gonads, and three-quarters of cell fates in worms with dorsal gonads. Specifically, our model was incapable of reproducing rows of 2° VPCs in the absence of 1° cells (e.g. 3° 3° 2° 2° 2° 3°) and asymmetric fate patterns in the neighbours of 1° cells (e.g. 3° 1° 2° 3° 3° 3°), both seen when the gonad is dorsally located.

To address these fate patterns, we introduce into our models changes in the behaviour of MAPK and the lateral

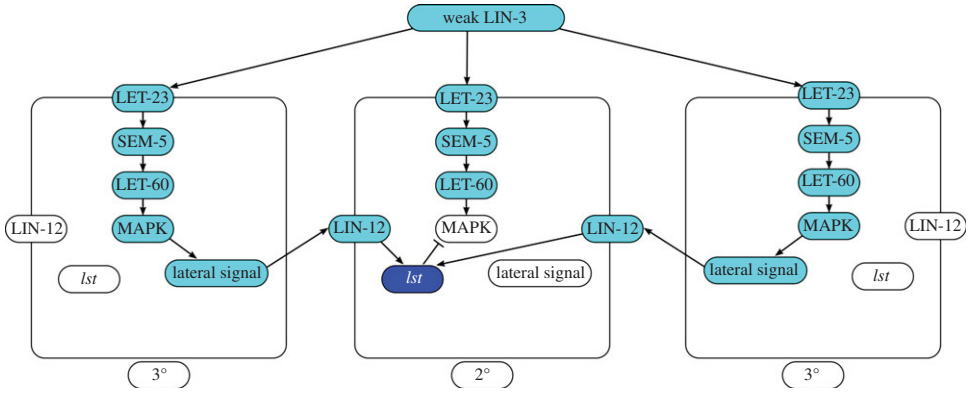

Figure 5. Effects of lateral summation. Active elements are indicated in blue, weakly active elements in cyan and inactive elements in white. Activating relationships are indicated by an arrow, inhibitory by a blocked line. A cell can experience a high lateral signal if it experiences two separate medium strength lateral signals. This can arise from a cell which produces no lateral signal neighbouring two cells which generate a medium lateral signal, or if a cell generates its own medium strength lateral signal and receives a medium strength signal from its neighbour.

Table 1. Experimentally observed fate patterns from the *dig-1* loss-of-function mutation (with both dorsally and ventrally located gonads). Cells are classified 1°, 2°, 3°, undivided (U) or abnormal (A). Taken from Thomas *et al.* [18].

P3.p	P4.p	P5.p	P6.p	P7.p	P8.p
3°	3°	2°	1°	2°	3°
3°	2°	1°	2°	3°	3°
U	1°	2°	3°	3°	3°
3°	1°	2°	3°	3°	3°
3°	A	2°	1°	2°	3°
3°	2°	1°	2°	2°	3°
A	2°	1°	2°	3°	3°
2°	1°	2°	A	2°	3°
2°	1°	2°	A	3°	3°
1°	2°	1°	2°	3°	3°
2°	2°	1°	3°	3°	3°
3°	3°	1°	3°	A	3°
3°	3°	2°	2°	2°	3°
2°	2°	2°	2°	3°	3°
3°	3°	3°	3°	3°	3°

signal. Rows of 2° cells have been observed in other systems such as *lin-12* gain-of-function mutations, but otherwise 2° cells are expected to be induced by a strong lateral signal sent by a neighbouring 1° cell. In wild-type animals, this lateral signal, generated by P6.p, activates LIN-12 in P5.p and P7.p, which induces the expression of *lst* genes and inhibits MAPK (lines 327–330, 417–423, 215–222 and 273–306, respectively, in EleganSDev.4ml). This inhibition is sufficient to overcome any weak MAPK activation resulting from the VPC experiencing a low or medium inductive signal. The activation of LIN-12 also prevents later activation of MAPK by strong inductive signals. To allow our model to reproduce fate patterns with rows of 2° cells, we propose that medium strength lateral signals, received from both neighbours simultaneously, can behave additively to induce a 2° cell in the absence of 1° cells (figure 5). This modification was, however, insufficient in isolation to correct the model. Early versions of our model which allowed additive behaviour showed cells which uniformly adopted the 3° fate, despite temporarily activating LIN-12. This occurred as in G1 the levels of LIN-12 activity rose, causing *lst* induction. In later stages of the cell cycle, LST reduced MAPK activity, which reduced the level of lateral signal being generated and leading to cells which had neither active LIN-12 nor active MAPK. To overcome this issue, we further modified the model to allow high levels of lateral signal to persist in the absence of LIN-3 stimulation. This persistence of lateral signal effects reflects the accumulation of cleaved, intracellular LIN-12 in the nucleus in G1, observed using GFP in both wild-type VPCs and VPCs arrested in G1 [9]. In addition to this, some rows of 2° cells include P3.p. To address this, we propose that both P3.p and P8.p are uniquely able to respond to their own soluble lateral signal component (lines 357–361 in EleganSDev.4ml). We justify this addition to the model as we expect neighbours of VPCs at other positions will trap received lateral signal, preventing autocrine signalling. By contrast, P3.p and P8.p are capable of generating lateral signal which is not captured by a neighbour, and thus are capable of responding to their own signal as it freely diffuses [18]. This behaviour may also suggest why worms, where all VPCs except P3.p are ablated, may adopt a 2° fate when LIN-3 expression is controlled externally, although rarely [34]. The lack of two VPC neighbours and resultant autocrine signalling may allow the cells to adopt the 2° fate, though the models as presented here do not allow us to explore this.

These new mechanisms are sufficient to explain fate patterns with rows of 2° cells observed in *dig-1* loss-of-function mutations, but do not explain how primary cells may exist with 2° and 3° neighbours (i.e. giving an asymmetric fate pattern). 1° cell fates can be induced without the VPC generating a strong lateral signal, as observed in *sem-5* or *lin-3* reduction-of-function mutations [14], but previous experimental data and models (with the exception of *dig-1* loss-of-function mutations) all indicate that the lateral signal is delivered with equal strength to both neighbouring VPCs. Some asymmetric patterns are possible without further model refinement. A 3° 3° 1° 2° 1° 3° pattern can arise from a medium strength LIN-3 signal arriving at cells P5.p and P7.p, leading to the cell at P6.p to experience a strong cumulative lateral signal and adopt a 2° fate. This would arise from medium levels of LIN-3 arriving at P5.p and P7.p, whereas the level of LIN-3 at P6.p was low or none (arising from the Brownian diffusion of limited quantities of LIN-3). The

model does not, however, allow for the generation of patterns such as 3° 1° 2° 3° 3° 3°. To overcome this issue, we have changed our model such that cells are able to respond to a low level of LIN-3. This level of LIN-3 is insufficient to generate a non-3° fate in isolation (line 432 in EleganSDev.4ml), but leads to the generation of a medium strength lateral signal (line 329 in EleganSDev.4ml). This allows the generation of asymmetric patterns by surrounding a 2° cell when one neighbour receives a medium level of LIN-3 (causing it to generate a medium strength lateral signal and assume a 1° fate) and the second neighbour receives a low level of LIN-3 (causing it to generate a medium strength lateral signal, and to adopt a 3° fate). The cumulative lateral signal received by the central VPC would be sufficient to induce a 2° fate and create the appearance of asymmetric induction of 2° fates. All cell fate patterns which can arise in our model are listed in the electronic supplementary material, table S3.

3.2. *let-23* mosaic mutations cause additional inductions by diffusion of unbound LIN-3

Genetic mosaic animals have been used successfully to study how fate patterns are dependent on the specific activation of distinct pathways, and to dissect the precise interactions by selectively inhibiting pathways in specific individual cells. This is controlled by creating a loss-of-function mutation of the protein of interest (i.e. *let-23*) and re-introducing protein expression through the use of an extrachromosomal array (typically with a marker to aid identification). Cells with this array express the protein normally, but during development the array is randomly lost in single cells during mitosis, creating clones of mutant cells. Mosaic analysis of *let-23* loss-of-function mutations has shown the importance of the lateral signal in establishing 2° fates in P5.p and P7.p, mosaic animals which lack LET-23 in all cells but P6.p are able to form the wild-type fate pattern [21,32,33]. In contrast, however, loss of LET-23 in P6.p leads to a variable fate pattern, most likely owing to the diffusion of unbound LIN-3 from P6.p.

Initial models without diffusion accurately predict the wild-type fate pattern in model animals with LET-23 expressed by P6.p. However, systems without LET-23 expression at P6.p only showed a single fate pattern, 3° 3° 1° 2° 1° 3°, which is rarely observed. Additionally observed patterns include worms with asymmetric fate patterns (i.e. 3° 3° 1° 2° 3° 3°) and wild-type-like, shifted patterns (i.e. 3° 2° 1° 2° 3° 3°). To consider asymmetric patterns, we propose that P7.p receives an amount of LIN-3 which varies across wild-type animals, between low and medium (which both give medium lateral signal, but in isolation lead to the 3° and 1° fates, respectively). This allows our model (including modifications built while modelling *dig-1* loss-of-function mutations) to accurately predict more observed phenotypes, without breaking the wild-type behaviour. This asymmetric pattern is reliant on the changes introduced in modelling *dig-1* loss-of-function behaviours, notably the creation of a level of MAPK activation which is sufficient to generate a lateral signal without leading to a 1° fate, further validating the modifications to the model.

To accurately model the remaining observed fate patterns, we included a notion of diffusion from cells lacking LET-23 but which experienced a high level of LIN-3. Cells which have high levels of LIN-3 delivered by the anchor cell (i.e. P6.p in otherwise wild-type animals) but which lack LET-23 are able to increase the level of inductive signal experienced by their neighbours (i.e. increase a medium strength signal to a high strength signal, lines 130–132 in EleganSDev.4ml). This modification allows our model to accurately predict observed fate patterns (electronic supplementary material, table S4). Additionally, the model also limits both the ability of LIN-3 to diffuse, and the absolute quantity of LIN-3 which is expressed. This is necessary as models which allow diffusion of LIN-3 from cells without a high level of LIN-3, or which allow LIN-3 to diffuse further than a single cell predict spurious fate patterns.

Finally, the subcellular location of LET-23 is explicitly included in the model as either basolateral or apical to reflect the trafficking of LET-23 to the basolateral membrane in wild-type cells by LIN-2/LIN-7/LIN-10. Loss-of-function mutations of LIN-2 cause LET-23 to be located in the apical membrane (lines 238–240 in EleganSDev.4ml).

3.3. Comparisons with historical data and altered fate patterns

Further validation of our models comes through some of the changes in fate patterns proposed by the new model, and through comparisons with older, previously unaddressed datasets. Several of the modifications to our model make changes which could be expected to change the predicted fate patterns arising from other mutations. Specifically, our new model predicts that *lin-12* loss-of-function mutations will occasionally show a fate pattern of 3° 3° 1° 1° 3° 3° in addition to the previously expected pattern of 3° 3° 1° 1° 1° 3°. This pattern arises specifically in worms where P7.p receives a low level of LIN-3. While this directly contradicts the consensus pattern selected in [16], examination of fate patterns observed in individual worms reveals that this pattern is mirrored in *lin-12* loss-of-function mutations [16]. An additional prediction from our model is that *lst* loss-of-function mutations will give both the 3° 3° 1° 1° 1° 3° and 3° 3° 1° 1° 2° 3° patterns (the latter arising from a low level of LIN-3 arriving at P7.p). This new pattern has also been observed in *dep-1* and *lip-1* double loss-of-function mutants, which inhibit LET-23 and MAPK, respectively [35]. The observation of non-1° cells at position P7.p in these two sets of mutations strongly supports the notion that the level of LIN-3 experienced at P7.p is variable between different worms, and that it is not necessarily sufficient to induce the 1° fate.

Furthermore, previous experimental work has shown that controlled, low levels of LIN-3 expressed by the VPCs are capable of inducing fate patterns with only 2° and 3° cells [34]. By controlling LIN-3 expression using a heat-shock promoter in worms lacking an anchor cell, it was found that low levels of LIN-3 could lead to fate patterns such as 3° 3° 2° 3° 3° 3°, and it was concluded that low LIN-3 levels were sufficient to cause cells to adopt 2° fates (in the absence of 1° fates). Our approach taken to model *dig-1* loss-of-function mutations allows us to create conditions which mimic the generation of low levels of LIN-3 by individual cells. Revisiting our model, we can observe that, through the creation of a low level of LIN-3 across three neighbouring VPCs, we are able to generate the pattern 3° 3° 2° 3° 3° 3°, reconciling these observations with the finding that the lateral signal and LIN-12 are responsible for cells adopting the 2° fate.

4. Discussion

Here, we have reported on the use of a novel abstraction for the modelling of biological systems. We have shown that logic programs are capable of encoding notions of temporal and spatial relationships between interoperating elements, while greatly reducing the computational overhead. We have demonstrated the validity of the model by applying it to model cell fate determination in the VPCs of *C. elegans*, and shown that we can reproduce previous results derived from both experiment and alternative modelling approaches. The increase in speed achieved by this novel abstraction has enabled us to massively search all possible mutant combinations, effectively offering a high-throughput technique for interrogating models. Alongside this, the specific nature of the abstraction offers us new ways to pose questions regarding non-deterministic behaviour induced by different mutants in the system, highlighting gaps in our understanding and allowing us to further refine our models. The changes in cell fate determination induced by *dig-1* loss-of-function mutations generate a large range of different patterns of fates in the VPCs. The ability of our approach to retrodict a set of conditions which lead a particular set of cell fates to arise uniquely enables us to ask whether our model is capable of generating an observed behaviour. Similarly, if certain patterns are known never to arise, then we can apply our retrodiction approach to confirm that such outcomes are unreachable. In future, we would expect to both expand our model to account for novel fate patterns (those not included here, in addition to mixed fates), as well as exploit retrodiction to classify novel mutants to suggest their mechanistic origins. The speed of model checking allows for rapid model refinement, and enables the use of large numbers of mutant combinations as a specification. As such, logic reasoning could be used in alternative biological systems alongside high-throughput genetic screens to rapidly build and validate models, allowing the user to postulate the genetic backgrounds linked to the experimentally determined phenotypes.

These new mechanisms provide novel insights into the determination of cell fate. While rows of 2° cells have been previously observed where LIN-12 is activated by mutation, there existed no previously established route to this observation. The concept of a persistent lateral signal supports the observation that the intracellular domain of LIN-12, generated by activation induced cleavage, accumulates in the nucleus in G1 and is only inhibited by MAPK activity through SUR-2. Finally, a low level production of lateral signal without inducing 1° fate allows a greater range of behaviours to be predicted and in combination with other rules is sufficient to describe all *dig-1* behaviours. The results of our modelling approach highlight a view of cell fate determination in *C. elegans* vulval development where a limited quantity of LIN-3 applied directly by the anchor cell to P6.p ensures 1° fate development, whereas the lateral signal from P6.p to P5.p and P7.p both determines 2° fate, but also prevents any free LIN-3 which does not get picked up by P6.p from determining the fate. Furthermore, this stresses the importance of irreversible timed events in cell fate determination, through the persistence of the lateral signal in weakly activated cells. Together, these new biological mechanisms (lateral signal summation, lateral signal persistence and stimulation of the lateral signal by low levels of LIN-3) reflect the powerful new insights that can be generated from modelling biological processes with logic programs. Alongside other formal approaches for understanding biological phenomena, we present logic programs as an important tool for the future development and exploration of executable biological models.

Acknowledgements. We thank Nir Piterman, Samin Ishtiaq and Aleksandra Watson for valuable discussions.

References

1. Fisher J, Henzinger TA. 2007 Executable cell biology. *Nat. Biotechnol.* **25**, 1239–1249. (doi:10.1038/nbt1356)
2. Sadot A, Fisher J, Barak D, Admanit Y, Stern MJ, Hubbard EJA, Harel D. 2008 Toward verified biological models. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **5**, 223–234. (doi:10.1109/TCBB.2007.1076)
3. Bonzanni N, Feenstra KA, Fokkink W, Krepska E. 2009 What can formal methods bring to systems biology? In *FM 2009: formal methods*, pp. 16–22. Berlin, Germany: Springer.
4. Setty Y, Cohen IR, Dor Y, Harel D. 2008 Four-dimensional realistic modeling of pancreatic organogenesis. *Proc. Natl Acad. Sci. USA* **105**, 20 374–20 379. (doi:10.1073/pnas.0808725105)
5. Efroni S, Harel D, Cohen IR. 2003 Toward rigorous comprehension of biological complexity: modeling, execution, and visualization of thymic T-cell maturation. *Genome Res.* **13**, 2485–2497. (doi:10.1101/gr.1215303)
6. Kam N, Cohen IR, Harel D. 2001 The immune system as a reactive system: modeling T cell activation with statecharts. In *Proc. IEEE Symp. on Human-centric Computing Languages and Environments*, pp. 15–22. Washington, DC: IEEE. (doi:10.1109/HCC.2001.995228)
7. Fisher J, Piterman N, Hajnal A, Henzinger TA. 2007 Predictive modeling of signaling crosstalk during *C. elegans* vulval development. *PLoS Comput. Biol.* **3**, e92. (doi:10.1371/journal.pcbi.0030092)
8. Fisher J, Piterman N, Hubbard EJA, Stern MJ, Harel D. 2005 Computational insights into *Caenorhabditis elegans* vulval development. *Proc. Natl Acad. Sci. USA* **102**, 1951–1956. (doi:10.1073/pnas.0409433102)
9. Nusser-Stein S, Beyer A, Rimann I, Adamczyk M, Piterman N, Hajnal A, Fisher J. 2012 Cell-cycle regulation of NOTCH signaling during *C. elegans* vulval development. *Mol. Syst. Biol.* **8**, 618. (doi:10.1038/msb.2012.51).
10. Shin Y-J, Nourani M. 2010 Statecharts for gene network modeling. *PLoS ONE* **5**, e9376. (doi:10.1371/journal.pone.0009376.g009)
11. Fioravanti F, Helmer-Citterich M, Nardelli E. 2012 Modeling gene regulatory network motifs using statecharts. *BMC Bioinform.* **13**, S20. (doi:10.1186/1471-2105-13-S4-S20)
12. Clarke EM, Grumberg O, Peled DA.1999 *Model checking*. Cambridge, MA: The MIT Press.
13. Clark A, Galpin V, Gilmore S, Guerriero ML, Hillston J. 2012 Formal methods for checking the consistency of biological models. *Adv. Exp. Med. Biol.* **736**, 461–475. (doi:10.1007/978-1-4419-7210-1_27)
14. Sternberg PW. 2005 Vulval development. In *WormBook*. See http://www.wormbook.org. (doi:10.1895/wormbook.1.6.1)
15. Gupta BP, Hanna-Rose W, Sternberg PW. 2012 Morphogenesis of the vulva and the vulval–uterine connection. In *WormBook* See http://www.wormbook.org. (doi:10.1895/wormbook.1.152.1)
16. Sternberg PW, Horvitz HR. 1989 The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in *C. elegans*. *Cell* **58**, 679–693. (doi:10.1016/0092-8674(89)90103-7)
17. Fisher J, Henzinger TA, Mateescu M, Piterman N. 2008 Bounded asynchrony: concurrency for

modeling cell–cell interactions. In *Formal methods in systems biology* (ed. J Fisher). Lecture Notes in Computer Science, vol. 5054, pp. 17–32. Berlin, Germany: Springer. (doi:10.1007/978-3-540-68413-8_2)

18. Chen N, Greenwald I. 2004 The lateral signal for LIN-12/Notch in *C. elegans* vulval development comprises redundant secreted and transmembrane DSL proteins. *Dev. Cell* **6**, 183–192. (doi:10.1016/S1534-5807(04)00021-8)
19. Sternberg PW. 1988 Lateral inhibition during vulval induction in *Caenorhabditis elegans*. *Nature* **335**, 551–554. (doi:10.1038/335551a0)
20. Thomas JH, Stern MJ, Horvitz HR. 1990 Cell interactions coordinate the development of the *C. elegans* egg-laying system. *Cell* **62**, 1041–1052. (doi:10.1016/0092-8674(90)90382-O)
21. Hajnal A, Whitfield CW, Kim SK. 1997 Inhibition of *Caenorhabditis elegans* vulval induction by gap-1 and by let-23 receptor tyrosine kinase. *Genes Dev.* **11**, 2715–2728. (doi:10.1101/gad.11.20.2715)
22. Jackson EK. 2013 FORMULA: formal modeling using logic programming and analysis. See http://research.microsoft.com/en-us/projects/formula/.
23. Jackson EK, Simko G, Sztipanovits J. 2013 Diversely enumerating system-level architectures. In *Proc. Int. Conf. on Embedded Software* (*EMSOFT*), pp. 1–10. Washington, DC: IEEE. (doi:10.1109/EMPSOFT.2013.6658589)
24. Jackson EK, Kang E, Dahlweid M, Seifert D, Santen T. 2010 Components, platforms and possibilities: towards generic automation for MDA. In *Proc. 10th ACM Int. Conf. on Embedded Software*, pp. 39–48. New York, NY: ACM. (doi:10.1145/1879021.1879027)
25. Jackson EK, Bjorner N, Schulte W. 2013 Open-world logic programs: a new foundation for formal specifications. Microsoft technical report MSR-TR-2013-55. See http://research.microsoft.com/apps/pubs/default.aspx?id=192963.
26. Gallese V, Goldman A. 2003 Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* **2**, 493–501. (doi:10.1016/S1364-6613(98)01262-5)
27. Pillai D, Sheppard E, Mitchell P. 2012 Can people guess what happened to others from their reactions? *PLoS ONE* **7**, e49859. (doi:10.1371/journal.pone.0049859)
28. Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, Heringa J. 2009 Executing multicellular differentiation: quantitative predictive modelling of *C. elegans* vulval development. *Bioinformatics* **25**, 2049–2056. (doi:10.1093/bioinformatics/btp355)
29. Giurumescu CA, Sternberg PW, Asthagiri AR. 2006 Intercellular coupling amplifies fate segregation during *Caenorhabditis elegans* vulval development. *Proc. Natl Acad. Sci. USA* **103**, 1331–1336. (doi:10.1073/pnas.0506476103)
30. Giurumescu CA, Sternberg PW, Asthagiri AR. 2009 Predicting phenotypic diversity and the underlying quantitative molecular transitions. *PLoS Comput. Biol.* **5**, e1000354. (doi:10.1371/journal.pcbi.1000354)
31. Li C, Nagasaki M, Ueno K, Miyano S. 2009 Simulation-based model checking approach to cell fate specification during *Caenorhabditis elegans* vulval development by hybrid functional Petri net with extension. *BMC Syst. Biol.* **3**, 42. (doi:10.1186/1752-0509-3-42)
32. Koga M, Ohshima Y. 1995 Mosaic analysis of the let-23 gene function in vulval induction of *Caenorhabditis elegans*. *Development* **121**, 2655–2666.
33. Simske JS, Kim SK. 1995 Sequential signalling during *Caenorhabditis elegans* vulval induction. *Nature* **375**, 142–146. (doi:10.1038/375142a0)
34. Katz WS, Hill RJ, Clandinin TR, Sternberg PW. 1995 Different levels of the *C. elegans* growth factor LIN-3 promote distinct vulval precursor fates. *Cell* **82**, 297–307. (doi:10.1016/0092-8674(95)90317-8)
35. Berset TA, Hoier EF, Hajnal A. 2005 The *C. elegans* homolog of the mammalian tumor suppressor Dep-1/Scc1 inhibits EGFR signaling to regulate binary cell fate decisions. *Genes Dev.* **19**, 1328–1340. (doi:10.1101/gad.333505)
36. Cui M, Chen J, Myers TR, Hwang BJ, Sternberg PW, Greenwald I, Han M. 2006 SynMuv genes redundantly inhibit lin-3/EGF expression to prevent inappropriate vulval induction in *C. elegans*. *Dev. Cell* **10**, 667–672. (doi:10.1016/j.devcel.2006.04.001)
37. Burket CT, Higgins CE, Hull LC, Berninsone PM, Ryder EF. 2006 The *C. elegans* gene dig-1 encodes a giant member of the immunoglobulin superfamily that promotes fasciculation of neuronal processes. *Dev. Biol.* **299**, 193–205. (doi:10.1016/j.ydbio.2006.07.019)
38. Bénard CY, Boyanov A, Hall DH, Hobert O. 2006 DIG-1, a novel giant protein, non-autonomously mediates maintenance of nervous system architecture. *Development* **133**, 3329–3340. (doi:10.1242/dev.02507)