A linear-encoding model explains the variability of the target morphology in regeneration

Daniel Lobo1, Mauricio Solano2, George A. Bubenik3 and Michael Levin1

1Department of Biology, Center for Regenerative and Developmental Biology, Tufts University, 200 Boston Avenue, Suite 4600, Medford, MA 02155, USA
2Cummings School of Veterinary Medicine, Tufts University, 200 Westboro Road, North Grafton, MA 01536, USA
3Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

A fundamental assumption of today’s molecular genetics paradigm is that complex morphology emerges from the combined activity of low-level processes involving proteins and nucleic acids. An inherent characteristic of such nonlinear encodings is the difficulty of creating the genetic and epigenetic information that will produce a given self-assembling complex morphology. This ‘inverse problem’ is vital not only for understanding the evolution, development and regeneration of bodyplans, but also for synthetic biology efforts that seek to engineer biological shapes. Importantly, the regenerative mechanisms in deer antlers, planarian worms and fiddler crabs can solve an inverse problem: their target morphology can be altered specifically and stably by injuries in particular locations. Here, we discuss the class of models that use pre-specified morphological goal states and propose the existence of a linear encoding of the target morphology, making the inverse problem easy for these organisms to solve. Indeed, many model organisms such as Drosophila, hydra and Xenopus also develop according to nonlinear encodings producing linear encodings of their final morphologies. We propose the development of testable models of regeneration regulation that combine emergence with a top-down specification of shape by linear encodings of target morphology, driving transformative applications in biomedicine and synthetic bioengineering.

1. Introduction

Large-scale morphology, including anatomy and patterning, is considered an emergent property of developing and regenerating organisms. There is no blueprint stored in the zygote; instead, a nonlinear encoding based on genetic and epigenetic networks drives development through the expression of diffusive [1] and reactive [2] biochemical signals [3–5], together with the mechanical and electrical properties of living cells [6–8]. Morphologies are high-level outcomes that unfold by the action of these networks that involve large numbers of concurrent low-level cellular mechanisms and their nonlinear interactions [9–13]. As in development, biological regeneration of organs, such as amputated amphibian limbs, involves the control of a complex network of genetic, biochemical and bioelectrical signals [14–17]. Indeed, many mechanisms necessary for regeneration are also present during development, and it is often stated that regeneration recapitulates morphogenesis [18,19]. Regeneration, therefore, is also commonly regarded as an emergent process controlled not only by a stored blueprint of the overall form, but also by nonlinear genetic encodings that control the action of low-level cellular mechanisms.

However, recent advances in developmental biology have revealed that, during development, low-level cellular mechanisms produce morphogenetic fields that prepattern the embryo; these serve as instructional information to which individual cells respond to form the resultant morphology [20–22]. These prepatterns are based on morphogen concentrations created by genetic networks and diffusion or reaction–diffusion mechanisms [23,24], electric
gradients created by electrical circuits formed within and between cells [21,25] or mechanical forces exerted and produced by the living tissue itself [6,26–28]. Thus, although formed by indirect low-level mechanisms during development, these fields and prepatterns represent a one-to-one encoding (a blueprint) from which further cellular mechanisms create the final morphology.

Moreover, the regenerating large-scale morphology of certain model organisms can be predictably altered, which suggests that the underlying mechanism of these regenerative processes is not based on a nonlinear encoding. As we review in the following sections, the target morphology—the shape to be restored during a regenerative process—of deer, planaria and fiddler crabs can be modified in a localized way through specific injuries or pharmacological treatments. The new regenerated morphology is either permanent or can last for several cycles of regeneration, without the need of reapplying the specific injuries or drugs that produced the change in the first place. Importantly, changing a nonlinear encoding to emergently regenerate a new shape or pattern represents a very hard inverse problem that cannot be efficiently solved [29], which discards the involvement of nonlinear genetic encodings in these regenerative systems. For example, given a genetic network (a nonlinear encoding) regulating the developing of a specific morphology, it is very difficult to determine what genes or links should be changed in order to produce a non-trivial desired specific change in the morphology, such as adding an ectopic limb or organ. A simple analogy can be made with ant behaviour. Each individual ant is following local rules about pheromone signals, and no single ant knows anything about the shape of the resulting anthill. Modelling the time evolution of such a system forward, it is easy to see how massively parallel execution of nonlinear rules can give rise to surprising and complex outcomes [30,31]. But, how would one modify the simple rules guiding each ant if one wanted the resulting anthill to have one extra lateral chimney?

This problem stands in sharpest focus in regenerative medicine, where we are faced with knowing which genes to tweak and how, in order to recreate a missing arm or an eye. While molecular pathways have made great strides in regulating the differentiation of stem cells into specific lineages, the incredible complexity of genetic and biophysical networks is a potent roadblock to the development of interventions that make desired changes at the level of anatomy (e.g. grow back the index finger, enlarge the lobe of one lung or rearrange craniofacial morphogenesis to repair a birth defect). A few examples of such anatomical change, leveraging developmental modularity, exist [32,33]. But, in general, the mathematics of nonlinear interactions in such complex emergent systems places fundamental constraints on our ability to know which gene products must be tweaked so that, when all cells carry out the resulting genetic network, a specific change of large-scale anatomy will result.

By contrast, in a system based on a linear encoding, the strategy would be different. For example, it is trivial to deduce from a one-to-one encoding (a blueprint, the simplest case of linear encoding) the changes necessary to specifically alter the morphology, because a change in the blueprint directly translates into the same change in the morphology. Knowing how the target morphology is linearly encoded in the chemical or physical properties of cells, one could change this information directly, and then rely on individual cells to build the shape without trying to micromanage the process [34]. Many issues of evolutionary developmental biology are impacted by the possibility that such linear encodings are used in embryogenesis. Moreover, the challenges of biomedicine for traumatic injuries and birth defects require that we take seriously models that may greatly augment our ability to direct growth and form at will. Finally, strategies for the bioengineering of novel hybrid structures in synthetic biology will be different depending on whether these linear encodings exist and can be manipulated.

While modern biology largely eschews anything that resembles the early theories of preformation, it must be remembered that regulative development, metamorphosis and regeneration have remarkable ability to *reach an anatomical goal state* despite considerable external perturbations of the number and locations of cells. Classical experiments [35] showed that early embryos can be divided or combined and give rise to perfectly normal animals [36]. During starvation, planarian flatworms continuously remodel and adjust organ sizes allometrically to precise proportions as available cell number is reduced [37]. In amphibian metamorphosis, artificial perturbation of tadpole facial anatomy becomes normalized into quite normal frog faces despite the fact that the organs start out in bizarre positions and must navigate around each other (in paths not predictable by evolution) to reach the correct frog face anatomy [38]. Tails grafted onto flanks of salamanders slowly remodel into limbs [39]. All of these are examples of cellular activity that is adaptively and flexibly controlled towards a target large-scale shape. An increasing subject of inquiry in genetic circuits seeks to show that emergent features of gene-regulatory networks include the systems property of robustness [40]. However, this has largely not been addressed at the level of large-scale shape [41], and there is a dearth of models to explain how cellular activity is guided towards the specific anatomical outcomes when the starting states were significantly different from normal (ruling out hardwired actions). One tempting set of concepts for investigating such models concern top-down [42–44] regulation (signals operating at the level of organ shape/size/identity, not cell behaviours), and implementation of algorithms that work towards specific goal states [45]. Such models often require the physical encoding of the target morphology.

In the next sections, we detail the target morphology variability exhibited by several organisms and discuss one-to-one and other linear encoding models that can explain this variability—a theme that has been out of favour for many years in the age of molecular cell biology. We show how these organisms are effectively solving an inverse problem—an achievement hardly possible with a nonlinear encoding, but trivial with a linear encoding. The experimental and theoretical evidence for the existence of a linear encoding of the regenerative target morphology suggests a rich and interesting research programme, which provides a necessary complement to the current roadmap for understanding self-assembly and repair of biological structures.

**2. Variable target morphology in regeneration**

The amputation of a salamander leg triggers a regenerative process combining growth and repatterning that restores the
original morphology [46,47]. As in most organisms with a regenerative capacity, the target morphology that this regenerative process creates is always the same: the original morphology of the wild-type limb. However, in some regenerative organisms, the target morphology that their regenerative process restores can be specifically altered through surgical manipulations or drugs. Among these animals are deer, planarian flatworms and fiddler crabs, whose characteristics are summarized in table 1 and detailed below. The most fundamental prediction of any linear encoding model is this: if a target morphology is linearly encoded and guides cell behaviour, then it should be possible to specifically change it, resulting in a stable change in the pattern to which the animal regenerates upon damage. This is indeed observed in a number of remarkable model systems.

### 2.1. Variable target morphology in antler regeneration

Antlers are deer appendages that cast and regenerate every year as extensions of the two permanent bony protuberances of the frontal bones called pedicles [48,49]. In general, only male deer grow antlers [50], following a cyclic process synchronized with the natural light cycle [51]. Initially, regenerating antlers contain a dense vasculature network and many sensory fibres that grow from the pedicle [52]. When growth stops, bone is formed in high quantities and the enveloping skin (velvet) dry and shed, leaving only the exposed solid bone [51]. Finally, the antlers are shed after the mating season, and a new cycle begins. The evolutionary adaptation of antler cyclic regeneration may be explained by the mechanical superiority of dry antler compared with wet bone in terms of elasticity, strength and impact absorption [53] and the difficulty for the body to maintain a junction between living and dead bone tissue [49].

Little is known about the control mechanisms of antler regeneration [50]. Stem cells located in a niche in the pedicle activate periodically, and are crucial for the regeneration of a new antler [50,54,55]. Hormones, such as testosterone and insulin-like growth factor I, are required for the growth of the pedicles and the development of antlers [50,51,56,57]. Research on local mechanisms of growth control has shown that retinoic acid, PTHrP/Indian-Hedgehog pathway, the canonical Wnt pathway and bone morphogenetic proteins are involved in the antler growth process [50]. Growing antlers are profusely innervated [58], and classical experiments have shown that electrical stimulation of the antler nerves during antler regeneration causes overgrowth and abnormal branching patterns [59–61]. Yet, the antler can regrow from a denervated pedicle, although smaller, lighter and with an altered shape [62,63].

The antlers’ shape is incomplete during the first years of life; until maturation, the number of branches and total length increase with age, where the morphological variability decreases. Because the morphology of the antler is species-specific, it is believed to be under control of genetic mechanisms [48]. However, experiments have shown that the antler target morphology can be specifically altered for several years owing to a single injury produced during regeneration—a phenomenon called *trophic memory* [65,66].

Figure 1 illustrates trophic memory in a white-tailed deer (*Odocoileus virginianus*) [56] and a Siberian wapiti (*Cervus elaphus xanthopygus*) [65]. Figure 1a shows three-dimensional reconstructions of computed tomography scans of the antlers of a white-tailed deer from year 5 to 8. The antlers regenerated normally in year 5 (first row), but, in year 6 (second row), an injury during the early developmental stages of antlerogenesis was suffered in the left antler. The injury altered the target morphology of this antler in that year, creating an atypical ‘royal’ (red arrow) instead of a single tine precisely in the location of the injury. This new target morphology was generated during years 7 and 8, producing a royal in the same location (green arrows) in the absence of any additional injury. In addition, the target morphology of the right antler was altered in a similar way, producing a royal in the reciprocal location during years 7 and 8 (blue arrows). Figure 1b shows the regenerated antlers (one side) of a Siberian wapiti during three consecutive years. During the first year shown, a slight cicatrize (red arrow) was produced by a cut off the dorsal portion of the germinative bud when the antler had reached nearly 40% of its normal length. Similar to the white-tailed deer, this injury altered the antler target morphology: the following 2 years, the regenerated antler presented a new tine at the site of the original injury (green arrows).

Trophic memory was also observed in fallow deer (*Dama dama*), red deer (*Cervus elaphus*) and moose (*Alces alces*) [65,66]. Stronger injuries, such a fracture in the pedicle, can cause stronger pattern alterations in the target morphology during the following regeneration cycles [59,66]. However, not all injuries produce trophic memory. For example, injuries near the end of the antler growth do not affect the antler development in the following cycles [65]. Remarkably, completely anaesthetized animals do not exhibit trophic memory either, regenerating the normal antler morphology during the following cycles after an injury [59,66], suggesting that some aspect of neural function [67,68] or bioelectrical communication among non-neural cells [69,70] is important for trophic memory to occur.

The implications of this phenomenon are profound for three reasons having to do with patterning information encoding in space and time. Spatially, the injury is made to a structure that is completely removed before next year’s growth shows an altered pattern. This reveals that the modification induced by the wounding was not a local event, but was transmitted a long distance to the growth zones at the scalp. Second, as with the other examples discussed below, this is a true example of a

<table>
<thead>
<tr>
<th>organism</th>
<th>regenerative part</th>
<th>target morphology</th>
<th>target morphology alteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>deer</td>
<td>antler</td>
<td>antler pattern</td>
<td>injury during regeneration</td>
</tr>
<tr>
<td>planaria</td>
<td>almost any body part</td>
<td>head, trunk, and tail regions pattern</td>
<td>amputation under GJC-blocking drugs</td>
</tr>
<tr>
<td>fiddler crab</td>
<td>chelipeds</td>
<td>handedness pattern</td>
<td>cheliped severance during development</td>
</tr>
</tbody>
</table>
kind of memory, because months pass between the original insult and the altered growth—whatever change has occurred, remaining cells must remember to alter the growth next year. Interestingly, related memory of positional information has now been demonstrated in salamander limb regeneration [71] and adult human fibroblasts [72,73]. Lastly, the ability to recreate an ectopic tine in the same place within a branched complex structure each year provides an ideal illustration of the inverse problem. Without a linear encoding, cells would be stuck with the intractable challenge of determining how to change their local growth rules so that next year, an ectopic tine was created in, and only in, the correct three-dimensional location. Although it is not yet known to what spatial accuracy the positional information is kept (what is the resolution of this memory system), the cut could have been made anywhere along the branched structure, rendering it very difficult to see how purely local growth rules could be altered to produce the needed ectopic growth in the right place. Such a phenomenon is not at all predicted by any emergent paradigm or molecular pathway model. By contrast, a linearly encoded target morphology model accommodates this finding easily, because once the linear representation of the branched structure is changed to include an extra tine, subsequent years’ growth will implement it. While this model system is relatively expensive, it is imperative to begin to investigate the mechanisms by which such branched morphologies can be stably encoded in tissue and the information altered by damage signals.

2.2. Variable target morphology in planaria regeneration

Planaria are flatworms with a complex bilaterally symmetric body plan, a brain allowing complex behaviours [74], and an outstanding regenerative capacity driven by a large adult stem cell population [75–77]. A cut planarian fragment as small as 1/279th can regenerate into a complete worm within one to two weeks [78]. Planarian regeneration involves the coordination of several mechanisms. After injury, the wound is closed with the help of muscle contraction [79], followed by the proliferation of a mass of new cells (called the blastema) at the injury site [80] counterbalanced by an increase in cell death (apoptosis) [81]. Regeneration
completes by a re-patterning of both the old and the new tissues, producing a new worm with all the parts adjusted to the new proportions for its now smaller size [82,83]. Many experiments have shown the existence of a carefully orchestrated communication between the new and old tissues necessary for the planarian regeneration [84–88]. These signalling mechanisms include the diffusion of morphogens [89], gap junctional communication [86,90,91], bioelectrical signals [92–94] and the nervous system [95]. However, despite the discovery of all these necessary mechanisms, no existing model can explain comprehensively more than one or two observed properties of planarian regeneration [75,96].

The target morphology of the bodyplan in planaria (the head, trunk and tail regions pattern) can be altered precisely and persistently through a combination of amputations and the blockage of gap junction communication via octanol in the medium [86], as illustrated in figure 2. Gap junctions are structures that allow current and small molecule signals to pass directly from the cytosol of one cell to that of a neighbour [97]—a system of physiological communication that plays important roles in pattern formation [90]. The planarian wild-type morphology consists of a head–trunk–tail polar pattern along the anterior–posterior axis (figure 2a). Amputated trunk fragments in a medium with octanol undergo a change in the target morphology (figure 2b), resulting in the growth of a head in both anterior and posterior wounds—producing two-headed bipolar worms (figure 2c). These changes in the target morphology are persistent: subsequent amputations regenerate the same altered morphology (figure 2d,e). This is the case even though the pharmacologically gap junction blocker that originally altered the target morphology (octanol) is washed out (as demonstrated by high performance liquid chromatography). The change in the target morphology is not mutagenic, because the octanol treatment does not change DNA [86] and its removal restores gap junctional communication very quickly [98].

These data highlight interesting new aspects of regeneration biology. First, the target morphology (the shape to which the animal regenerates upon damage) is stably altered by a treatment that perturbs real-time physiological signalling but does not impact the animal’s genomic sequence. Second, this radical change of bodyplan and behaviour is stable with respect to the animal’s normal mode of reproduction (splitting followed by regeneration), raising the possibility that such physiological changes might play a role during evolution [99]. Indeed, if such worms survived in the wild, then future scientists encountering the one-headed and two-headed worms in a pond might be tempted to sequence their genomes in a search for the speciation event. The failure of this strategy serves as a reminder that not all patterning information is present at the genetic level in an adult organism. One is immediately tempted to suggest epigenetics as a mechanism [100]; chromatin modification may certainly be involved; however, the key here is that it is not sufficient. The posterior-facing (tail) wound cells that are reprogrammed to build a head may indeed be epigenetically altered by temporary changes in gap junction-mediated signals, but this tissue is removed in subsequent cuts! The worm that regenerates as a two-headed animal in future rounds of regeneration is made from a fragment that initially is anatomically normal mid-trunk tissue. Thus, whatever the nature of the altered target morphology memory (epigenetic, bioelectrical or otherwise), it is distributed throughout the animal and not local—even trunk tissue knows that if damaged, then it needs to make a worm with two heads.

We are currently working on formulating and testing global models of target morphology storage in bioelectrical networks of non-neural somatic cells.

2.3. Variable target morphology in fiddler crab regeneration

Adult male fiddler crabs (Uca lactea) possess two asymmetrical chelipeds with different size: the major chela (crusher)
used for aggressive and courtship displays, and the minor (cutter) used for prey capture and grooming [101,102]. Like many crustaceans, fiddler crabs can sever their own limbs (autotomy reflex), after which they can regenerate a new appendage [103]. In contrast to shrimps, lobsters and other crabs, the adult fiddler crab has a permanent handedness (left or right, with equal probability [104]), which is not genetically determined, but attained during the early years of development [104,105].

Hormones have an important role in the regulation of moulting and limb regeneration in crabs [106,107]. RNAi-mediated silencing of the genes encoding the ecdysteroid steroid hormone receptors, EcR/RXR, arrests blastema formation and inhibits the regeneration of functional limbs [108]. However, no physiological mechanism is known for the control and maintenance of the handedness in the fiddler crab, although a dynamical mathematical model has been suggested [109].

Figure 3 illustrates the acquisition and maintenance of the fiddler crab handedness, that is, the establishment of its target morphology. Crabs develop two small chelipeds of equal size (no handedness; figure 3a), but, during this early stage, a cheliped is naturally lost (both sides with equal probability [104]; figure 3b). After losing a cheliped, the remaining cheliped then develops into the giant size, whereas the lost cheliped is regenerated to the original small size (figure 3c). This first amputation of a cheliped establishes permanently the target morphology in the crab. Further amputations of the giant, the normal or both chelipeds result in the regeneration of the same morphology, that is, the same handedness established with the first amputation.

The acquisition of handedness in the fiddler crab not only represents another example of patterning information not encoded at the genetic level, but also reveals a new mechanism to establish the target morphology in regeneration. In contrast to the deer antler and planaria, fiddler crabs do not encode an innate target morphology. Instead, during development, the target morphology (handedness) is established according to a random event: the side in which a cheliped is lost. Once this event has occurred, any further regeneration follows this established target morphology, becoming impossible for the crab to develop a different handedness. Moreover, even after amputating both chelipeds—which implies starting with the same morphology in both right- and left-handed crabs—the same established handedness regenerates. Hence, the physical encoding of the target morphology must be located not in the giant or normal cheliped, but somewhere in the crab body. Still, no experimental procedure has been found to alter the target morphology of the crab once it has been established. Finding such manipulations would shed light on the mechanisms responsible for maintaining the target morphology in these organisms.
3. Types of morphological encodings and the inverse problem

3.1. Linear and nonlinear morphological encodings

An organism that is able to develop or regenerate a body part needs to produce growth consistent with the appropriate morphology of that body part. For example, the antler morphology, the planarian bodyplan (head–trunk–tail polarity) and fiddler crab handedness information need to be stored within the organism in order to regenerate these morphologies. We can distinguish two types of morphological encoding according to the type of function necessary to decode the encoding: linear and nonlinear encodings.

A linear encoding is based on a linear mapping between elements of the encoding and elements of the outcome. The simplest linear encodings are the one-to-one encodings (also called direct encodings). Similar to a blueprint, a one-to-one encoding is formally a bijection: every element of the encoding is paired with exactly one element of the outcome, and every element of the outcome is paired with exactly one element of the encoding. During development, many organisms follow an isometric (same scale) one-to-one encoding, which are usually referred to as prepatterns. For example, the early striped prepatterns in Drosophila represent an isometric one-to-one encoding of the future morphology of the larva: every stripe of high protein concentration in the embryo corresponds to a specific segment in the larva and vice versa. More complex linear encodings are based on linear maps between elements of the encoding and elements of the outcome: a linear transformation produces the outcome according to the encoding. There are many uses of linear encodings in engineering. For example, a very simple method for image compression is run-length encoding, where sequences of the same data value are stored as a single data value and count, instead of the original run. In this way, the line of pixels ‘WWWWWBBBB’ (where ‘W’ represents a white pixel, and ‘B’ a black pixel) can be encoded with the shorter string ‘5W4B’. Hence, a symbol in a linear encoding can correspond to several locally related symbols in the outcome.

On the other hand, a nonlinear encoding (also called indirect encoding) is based on iterative methods and interconnected components, where a specific element in the code does not correspond to a specific element in the outcome. Many developmental systems are based on nonlinear encodings, because a genetic network is a nonlinear encoding of the developing morphology. For example, any branching structure (such as vascular system, lung, kidney, liver, etc.) is not linearly encoded branch by branch with individual genes, but in a regulatory network that produces the emergent branching pattern through biochemical interactions [111]. The gene-regulatory network together with the local laws of physics and chemistry govern such dynamical emergent systems. In this way, the outcome is grown (or regenerated) according to the set of rules and interactions orchestrated by the encoding.

Figure 4 illustrates the differences between a one-to-one, a linear and a nonlinear encoding for storing an artificial branching morphology. The one-to-one encoding (figure 4a) consists of a blueprint (blue) of the final morphology (green); every element in the encoding corresponds to an element in the morphology and vice versa. The general linear encoding (figure 4b) is based on a string of symbols (blue) that are interpreted according to turtle geometry [112] to produce the final morphology (green): a ‘turtle’ that leaves a trace moves according to the sequence of symbols read in the string. The symbol ‘F’ advances the turtle in a straight line (creating a segment in the morphology), the symbols ‘+’ and ‘−’ increase or decrease respectively the angle that the turtle is facing, and the brackets create a turtle subpath from the current position. In this encoding, a single symbol ‘F’ corresponds to all the constituting parts that form a straight segment of certain length, while other elements (such as brackets) have no direct correspondence with any particular element in the morphology. Finally, the nonlinear encoding system (figure 4c) adds an extra layer of complexity: a parallel rewriting grammar (L-system, much used for modelling biological development [113,114])
generates the string for the turtle geometry. The encoding is a single short rule (blue), which is applied iteratively, replacing the left-part of the rule (the symbol ‘F’) with the right-part of the rule (the string ‘F+[F−F][F]’). Starting with the string ‘F’, the rule is applied three times to obtain a final string, which is used by the turtle geometry to generate the final morphology (green). Note that, owing to the iterative process, there is no specific relation between an element of the encoding and an element of the final morphology: a specific ‘F’ symbol in the rule does not represent a specific segment in the morphology.

One-to-one, linear and nonlinear encodings differ in many properties. Nonlinear encodings can achieve great efficiency when encoding repetitive patterns (with the expense of higher computational time): figure 4c shows how a very short rule encodes a complex branching pattern (which can continue to grow in size without needing a longer rule). By contrast, in a one-to-one encoding (figure 4d), repeated parts in the outcome are encoded with repeated parts in the encoding, making the encoding spatially inefficient. A linear encoding (figure 4b) represents a balance between spatial efficiency and computational time. In addition, nonlinear encodings have valuable properties in an evolutionary context, facilitating the evolutionary emergence of modularity, scalability, adaptability, novelty and diversity with respect to a nonlinear encoding [115–118]—yet, the combination of nonlinear and linear encodings can outperform any of the two alone [119]. However, while a linear encoding for a given morphology is very easy to produce, it represents a very difficult problem for a nonlinear encoding. This characteristic can be formalized as an inverse problem.

### 3.2. Forward and inverse problems

The main goal of developmental biology is to explain and predict the shape that will result from a given state of an egg or embryo. The main goal of regenerative medicine and synthetic bioengineering is to learn to provide perturbations to change the course of the complex patterning system to result in desired changes in morphology (e.g. to induce stem cell derivatives to grow an eye or limb). In order to facilitate a mathematical study, the processes of biological development and variable regeneration can be considered a forward and an inverse problem, respectively. In general, a developmental or regenerative process can be represented with the following mathematical expression: \( G(m) = d \), where \( G \) is the operator representing the invariable physical mechanisms that transform the parameters \( m \), the code (including environmental factors) for the target morphology, into the output \( d \), the developed or regenerated morphology. Using this abstraction, we can now formalize the forward and inverse problems.

The forward problem consists of finding the output \( d \) that is produced from certain given operator \( G \) and parameters \( m \). From this definition, it is clear that biological development solves a forward problem: the morphology of the developed organism (output \( d \)) is obtained from the invariable physical mechanisms (operator \( G \)) combined with the genetic code and epigenetic information stored in the zygote (parameters \( m \)).

By contrast, an inverse problem consists of finding the parameters \( m \) that produce a given output \( d \) with an operator \( G \). The variable target morphology in regeneration discussed in the previous section is an example of an inverse problem: in order to specifically alter the regenerative target morphology, it is necessary to find a new code (parameters \( m \)) that produces such new morphology (output \( d \)), using the invariable physical mechanisms (operator \( G \)). For example, an inverse problem would be to create a genome that would produce a starfish-shaped creature with an elephant-like foot below and vertebrate eyes at the tips of each arm.

It is worth noting that the inverse problem has been studied in many scientific fields using different terminologies [120]. In computer science, \( G \) may be called an algorithm, a program, a procedure or the rules of a machine; \( m \) may be called the input, the arguments or the input variables; and \( d \) is usually called the output. An inverse problem in computer science consists of finding the input that produces a specific output for a given algorithm. In mathematics, \( G \) may be called a function or an equation; \( m \) may be called the input or the arguments; and \( d \) may be called the output or the value returned by a function. An inverse problem in mathematics consists of finding the argument for a specific function to return a given value. In physics, \( G \) may be called a model or a formula; \( m \) may be called the parameters, the independent variables or the input signal; and \( d \) may be called the result, the dependent variables or the output signal. An inverse problem in physics consists of finding the parameters that produce a specific result in a given model.

The organisms with a variable target morphology presented in the previous sections solve an inverse problem. When a developing antler is injured, its morphology is altered with a new royal precisely at the site of the injury. Furthermore, the encoding of the antler morphology is also altered to produce this new morphology in the following regenerative cycles (figure 1). Creating this new code for the new antler morphology is equivalent to solving an inverse problem, because the local code and rules governing cell behaviour have to be altered in precisely the right way to result in this new large-scale shape. Likewise, a temporary inhibition of gap junctional communication in planaria permanently alters the encoding of the target morphology, producing two-headed animals after each cut (figure 2). Again, creating the new encoding for the altered target morphology requires solving an inverse problem: what different rules will trunk cells follow if, in the future, they are surgically isolated and called upon to build an entire worm, which has to be two-headed? Finally, in the case of the fiddler crab, the encoding of the target morphology is initially created after losing a cheliped (figure 3), also solving an inverse problem of morphogenesis.

### 3.3. The forward problem is easy with linear and nonlinear encodings

The forward problem, obtaining the output given the operator and parameters, is easy to solve with both linear and nonlinear encodings. In a linear encoding (including one-to-one encodings), a simple algorithm applied to the encoding transforms it into the output (figure 4a,b). In the case of a nonlinear encoding, the output is growth applying the operator to the parameters. Starting with a simple state (the zygote), a nonlinear encoding orchestrates the growth of the resultant morphology (figure 4c). The forward problem, therefore, is easy for both types of encodings; however, we will find important differences in the case of the inverse problem.
3.4. The inverse problem is easy with linear encodings, but hard with nonlinear encodings

Solving the inverse problem—finding the specific code that grows a given output—is straightforward with a linear encoding. Because all linear functions are invertible, we can apply the inverse function of a particular encoding to any given output to obtain its corresponding code. In the case of one-to-one encodings, it is trivial to create a blueprint from an existing building: for every element in the building, the corresponding symbol is added to the blueprint. Likewise, it is easy to know how to add another room onto such a building: draw (or copy) such a room onto the plan, precisely where it is to go, and the apparatus that interprets the plan will implement the change. The inverse problem with linear encodings such as the run-length is equally easy: the product ‘WWWWWBBB’ can be easily transformed into the code ‘5W4B’. Indeed, the inverse problem using a linear encoding is a well-posed problem, and efficient analytical solutions can solve it. Thus, with linear morphological encodings, biologists are freed from a limit imposed by nonlinear mathematics—their task reduces to finding the mechanisms by which pattern and form are encoded in properties of tissue and by which cells interact with this information to guide their local activity.

By contrast, solving the inverse problem is very hard in the case of a nonlinear encoding. The algorithm that transforms a code into a product cannot be applied in reverse with a nonlinear encoding, because there is no reversible relation in general between output and code. For example, it is very hard to create a genome (nonlinear encoding) that produces a given morphology, because there is no linear mapping between an element of the morphology and an element of the genome. The inverse problem using a nonlinear encoding is not a well-posed problem, because there is no analytical solution to find the inverse of any nonlinear function.

Figure 5 illustrates the key difference between the linear and nonlinear encodings that makes solving the inverse problem easy or hard, respectively. With a linear encoding, a small change in the code produces a local small change in the output; but, with a nonlinear encoding, a small change in the code can produce a large change in the output. To illustrate this important difference, we computed all the possible morphologies that result from inserting a single short substring ‘[+F]’ or ‘[−F]’ (which by itself encodes a single segment) in all possible locations of the linear and nonlinear codes (strings) presented in figure 4. Figure 5a shows the resultant morphologies in the case of the linear encoding: inserting either of the substrings results in the addition of a single new tine exactly at the location of the insertion. By contrast, figure 5b shows the resultant morphologies in the case of the nonlinear encoding: inserting either of the simple substrings causes the development of many new branches with no direct relation between the location of the insertion and the location of the change. The large and delocalized change in the output is due to the recursive process characteristic of nonlinear encodings and the lack of a reversible mapping between encoding and product. It is clear now why a linear encoding can solve the inverse problem easily: to encode a new morphology with an extra tine, only a new short substring in the corresponding location of the new tine is necessary. Furthermore, with a linear encoding, all possible morphologies with an extra tine can be obtained by inserting the new substring in the corresponding place. By contrast, with a nonlinear encoding, it is not possible to obtain all possible morphologies in general, and there is no clear way to change the code to produce a given output.

The inverse problem with nonlinear encodings is pervasive in many scientific fields, and there is not a simple method to solve it in general. In the computer science field, it was formally demonstrated to be impossible to build a program (a Turing machine, which can be considered as a nonlinear encoding) to solve the general inverse problem in a finite time [121]. In a mathematical sense, inverse problems are from the class of problems where the input is not a continuous map to the output [29]. There is no general analytical solution for finding the rules that indirectly generate a given output [122], stochastic and heuristic search methods are the usual approaches to find approximate solutions for complex inverse problems [120,123–127]. While the inverse problem of producing genomes for morphologies adapted to their environments is solved by nature via the stochastic process of evolution, such
Figure 6. Many organisms develop according to a nonlinear encoding producing a one-to-one linear encoding of the final morphology. (a) In Drosophila, the maternal gene factors and a complex gene network including gap, pair-rule, and segment polarity genes form together a nonlinear encoding of the homeotic gene expression pattern that is produced in the larva embryo. This expression pattern is a blueprint, a one-to-one encoding, of the final morphology of the fly: each part of the fly morphology is determined by the expression of a homeotic selector gene. (b) In hydra, a reaction–diffusion mechanism (nonlinear encoding) produces a concentration prepattern (one-to-one encoding) of the HyAlx gene, which establishes the location where the hydra tentacles will grow. (c) In Xenopus, a bioelectric network (nonlinear encoding) produces an electrical prepattern (one-to-one encoding) of the tadpole face morphology: ectoderm regions with hyperpolarized cells (brighter) establish the developmental location of the eyes (blue marks) and mouth (red mark). Embryo and adult Drosophila cartoons adapted from [128]. Hydra pictures adapted from [129]. Electric frog embryo picture adapted from [130].

strategies are not good candidates for biological mechanisms of regenerative shape change because they operate far too slowly to allow real-time morphogenesis. Instead, we propose the existence of linear encodings to guide regeneration, which explains the variable target morphology showed in the model organisms presented above.

4. Linear encodings in development and regeneration

4.1. Two-step nonlinear – linear encodings in development

Many organisms follow a two-step process during development, combining a nonlinear with a one-to-one linear encoding. Figure 6 shows three examples of this two-step mechanism. During Drosophila development (figure 6a), the maternal effect genes (such as bicoid and nanos) and a complex genetic network (including gap, pair-rule and segment polarity genes) together represent a nonlinear encoding of the patterns of the homeotic gene expressions, which emerge from the genetic interaction, diffusion and reaction of such gene products [33,131 –136]. However, the homeotic gene expression pattern represents a one-to-one linear encoding for the subsequent morphology of the fly: the location of the expression of every gene product specifies the location of a specific fly body part. A second example can be found in hydra (figure 6b), where a reaction–diffusion process [137] (a nonlinear encoding) creates a characteristic expression pattern of the HyAlx gene [129]—long-range pattern emerges from the chemistry of local rules. Then, this pattern serves as a one-to-one encoding that establishes a blueprint with the locations where the hydra tentacles will grow. Another example can be found in Xenopus development (figure 6c), where the bioelectric networks created by ion channels and gap junctions (nonlinear encoding) produce patterns of cells with different resting potentials [130]. These membrane voltage patterns (one-to-one linear encoding) establish the morphology of the tadpole face (figure 6c), including the location of the eyes (blue areas in the figure) and mouth (red area).

Similar to the artificial experiments showed in figure 5, experimentally altering the nonlinear or linear encodings in these biological systems results in global or localized changes in the morphology, respectively. In Drosophila, landmark experiments showed how mutations in the gap genes (nonlinear encoding) globally affect the development of many regions in the larva [131]. By contrast, mutations in the homeotic genes producing the one-to-one linear encoding pattern result in localized changes in the developed fly, such as transforming the morphological identity of one specific segment into another [33,138]. In hydra, altering the genetic and reaction–diffusion chemical networks (nonlinear encoding) can produce global changes in morphology, such as phenotypes with ectopic tentacles all over the body [139,140]. Similarly, alterations of the membrane voltage regulation of cells (nonlinear encoding) during Xenopus development cause global malformations in craniofacial morphogenesis [130]. Remarkably, altering the membrane voltage of a small group of non-eye cells anywhere in the body, that is, directly manipulating the membrane voltage field (one-to-one linear encoding), induces the development of
4.2. A linear encoding can explain the variable target morphology in regeneration

Deer antlers, planaria and fiddler crabs can alter their encoded target morphology in a precise and lasting manner, for which an inverse problem needs to be solved by the cells that must rebuild each structure. In order to change the target morphology, they must know the local actions that will produce the new morphology. However, we have seen that solving the inverse problem with a nonlinear encoding is very hard. We propose, therefore, the existence of a linear encoding of the target morphology in these regenerative organisms, which can explain the variability of their regenerative morphologies. More broadly, we argue for a greater consideration of linear target morphology models in developmental and synthetic biology: the community must consider and test not only emergent nonlinear models popular in systems biology and complexity science, but also models that postulate an explicit encoding of target shape.

As in the developmental systems of Drosophila, hydra and Xenopus shown in figure 6, deer antlers and planaria may use a two-step process during development, combining an initial nonlinear with a lasting linear encoding. A transcriptional network encoded in the genome represents a nonlinear encoding, which produces a linear encoding, represented by a list of sequential instructions (antler) or a blueprint (planaria) of the target morphology to develop. This linear encoding can readily orchestrate the regeneration of cast antlers, or amputated planarian body parts. Importantly, this mechanism can explain the variable target morphology present in these organisms. Figure 7a illustrates this mechanism. A transcriptional network encoded in the genome represents a nonlinear encoding, which produces a linear encoding, represented by a list of sequential instructions (antler) or a blueprint (planaria) of the target morphology to develop. This linear encoding can readily orchestrate the regeneration of cast antlers, or amputated planarian body parts. Importantly, this mechanism can explain the variable target morphology present in these organisms.

4.2.1. Injuries or drugs can alter locally the linear encoding of the target morphology.

As in the developmental systems of Drosophila, hydra and Xenopus shown in figure 6, deer antlers and planaria may use a two-step process during development, combining an initial nonlinear with a lasting linear encoding. A transcriptional network encoded in the genome represents a nonlinear encoding, which produces a linear encoding, represented by a list of sequential instructions (antler) or a blueprint (planaria) of the target morphology to develop. This linear encoding can readily orchestrate the regeneration of cast antlers, or amputated planarian body parts. Importantly, this mechanism can explain the variable target morphology present in these organisms. Figure 7a illustrates this mechanism. A transcriptional network encoded in the genome represents a nonlinear encoding, which produces a linear encoding, represented by a list of sequential instructions (antler) or a blueprint (planaria) of the target morphology to develop. This linear encoding can readily orchestrate the regeneration of cast antlers, or amputated planarian body parts. Importantly, this mechanism can explain the variable target morphology present in these organisms.
Importantly, not all linear encodings need to be isometric (like a prepattern), but they can exist on a different scale from the morphologies that they encode. In the same way that an architect blueprint or compressed image has a smaller size than the building it represents, a linear encoding can be physically smaller than the morphology it encodes. In particular, the linear encoding in deer antlers, planaria and fiddler crabs cannot be a one-to-one prepattern as illustrated in *Drosophila* development (figure 6a). The linear encoding of the deer antlers must reside in a smaller scale outside of the antlers themselves, because they are cast every year and grown anew from the pedicles. In wild-type planaria, amputating a piece of the worm does not alter its target morphology, meaning that it should be encoded redundantly over their bodies in a smaller scale. Similarly, chelipeds with the correct size are developed even from crabs with both chelipeds amputated. A difference in scale between the encoding and the morphology implies the existence of a method to generate the scaled-up morphology form the linear encoding. Examples of such methods include means-end algorithms that specifically consult the linear encoding to perform the actions necessary to reach the right morphology from the current state. Notably, these algorithms predict the extraordinary capacity during embryonic development to fix perturbed morphologies [38].

Many biological mechanisms can store a linear encoding of the target morphology. The trophic memory in antlers has been suggested to be physically located in the deer nervous system, possibly in the brain [65,66]. Although a neural network is powerful enough to store nonlinear encodings, it is also possible for it to store any less complex encodings, such as the linear encodings that we propose here. Moreover, because the antlers are innervated, the trauma information can be sent through the neurons located in the antlers, which can precisely alter the encoded target morphology (solving the inverse problem), but only if this encoding is linear. Interestingly, recent data using morphometrics and laser ablation in *Xenopus* larval tail model revealed that the central nervous system far away from the wound site seems to carry instructive information for shape of the regenerated appendage [142,143]. Other ways of storing morphological information with a linear encoding include chemical maps maintained by feedback loops (such as the transcriptional memory of *Hox* genes [144]), physical structures (such as nerve cords) or electrophysiological mechanisms based on gap junctions (as suggested for cardiac memory [145,146]).

### 5. Conclusion

Deer, planaria and fiddler crabs possess the capacity to alter the target morphology achieved through regeneration, a phenomenon that requires altering the individual behaviours of thousands or millions of cells to achieve a large-scale anatomical goal state. If we accept the common view that living systems are fundamentally computational in nature [147–149], then these organisms necessarily use a linear encoding for their target morphology because they are able to alter it precisely in the location of specific injuries and, hence, solve an inverse problem. We have argued in this paper that a combination of a nonlinear and a linear encoding can explain the variable target morphology during regeneration. While a nonlinear encoding can facilitate the evolution of modularity and diversity, a linear encoding may be an efficient mechanism to facilitate regeneration.

Nonlinear and linear encodings differ with respect to exactly what is specified (encoded) in the physical medium, and how direct (how much decoding) needs to take place to derive the target morphology. Nonlinear encodings specify recursive rules for cell or molecule behaviour: the morphology emerges as a result of applying the rules. Examples of nonlinear encodings include gene-regulatory networks specifying cell interaction rules and cellular automata such as the game of life. On the other hand, linear encodings specify simple rules (a linear transformation) that need to be applied to the code to obtain the final morphology. Prepatterns are the simplest type of linear encodings: a one-to-one map exists between the code and the target morphology. There are many examples of prepatterns in developmental model organisms, such as the *Hox* gene gradients in *Drosophila* that directly specify axial identity. However, not all linear encodings are prepatterns; other linear encodings can bear little or no direct relationship to the final shape. Examples of this more complex type of linear encoding include simple algorithms for compressing images and turtle graphics methods to generate complex shapes. The differences between these encodings are summarized in table 2.

The fundamental difference between nonlinear and linear encoding models suggests the need to expand the current approach in regenerative biology from an exclusive focus on genetic networks and pathways (nonlinear encodings), to take into account models based on spatial encodings of morphological information. Importantly, the linear encoding of the target morphology’s representation in tissue properties can be very straightforward (such as *Hox* gene prepatterns that specify underlying tissue fate), or they can be encoded in a more complex manner (linear transformations). For example, a neural network (or a bioelectrical network of non-neural cells [69,150]) could store information that guides growth towards a specifically remembered shape, but the information is not stored in a simple ‘image’ of the final product but in the distribution of node activation strengths. Either type of encoding can function as target end states of cybernetic goal-seeking mechanisms, such as algorithms that form a shape by comparing the current state with a target state—a novel approach to model a range of regenerative phenomena as discussed above.

Thus, our proposal differs from existing models of chemical prepattern in these two critical ways: (i) we propose that

<table>
<thead>
<tr>
<th>Encoding type</th>
<th>Code</th>
<th>Decoding process</th>
<th>Forward problem</th>
<th>Inverse problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinear</td>
<td>Genetic networks, recursive rules</td>
<td>Recursive mechanism</td>
<td>Easy</td>
<td>Hard</td>
</tr>
<tr>
<td>Linear</td>
<td>Simple rules</td>
<td>Forward mechanism</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>One-to-one</td>
<td>Prepattern, blueprint</td>
<td>Replacing map, scaling</td>
<td>Easy</td>
<td>Easy</td>
</tr>
</tbody>
</table>
target morphology can be explicitly encoded in cell properties far more complex than gradients and prepatterns corresponding to spatially overlying tissue fate and (ii) this information, owing to its linear nature, could be still explicitly ‘read’ by processes seeking to repair and remodel shape, and ‘written’ by processes that alter the pattern to which future growth should conform.

Indeed, models based on a linear encoding of the target morphology have unique testable implications that are not predicted by any existing emergent genetic model. A linear encoding predicts the capability to experimentally produce precise and lasting morphological alterations during the lifetime of a single organism, as we have discussed for deer antlers, planaria and fiddler crabs. Indeed, planaria are some of the most plastic model organisms in regenerative biology, with more than 25 known experimental phenotypes, as recorded in the planarian phenotype database Planto[151]. By contrast, nonlinear encoding models can account only for a much reduced phenotype landscape (those phenotypes that can be produced with a specific set of rules) [119]. Moreover, a linear encoding mechanism can readily explain why cancerous cells grafted in developing embryos or regenerating organs stop their neoplastic behaviour and become integrated as normal tissue [152–156]. A linear encoding mechanism during development and regeneration directly contains spatial information of the target morphology, which can reprogram the cancerous cells according to their location to become part of the encoded target morphology. Linear encoding models thus suggest alternative approaches to cancer normalisation (in line with the views of cancer as a problem of tissue organization [157,158]) focused on activating cellular responses to fields of non-local patterning information [21,34,159,160].

In addition, the inherent plasticity of linear encodings is of exceptional importance in regenerative biomedicine. Finding the changes in a nonlinear encoding necessary to emergently restore a desired morphology (solving the inverse problem) is a very complex task and currently regarded as a long-term goal in regenerative medicine. By contrast, properly altering a linear encoding to produce a specific morphology is a much easier task. This was demonstrated with the electric map during Xenopus development: an ectopic eye can be experimentally induced in any desired location by changing the transmembrane voltage levels (linear encoding) present in the embryo precisely in that location [141]. Cells were coaxied to implement a complex organ without the need for the experimenter to micromanage the progress. Thus, the discovery of morphological linear encodings would pave the way for novel medical procedures to regenerate amputated body parts long before we had the capability of building such a complex structure directly or of altering gene regulatory networks to make the needed change and no more.

Finally, this new perspective on regeneration can also benefit the engineering and computational fields, inspiring novel mechanisms for resilient self-assembly robotics [161–163] and novel heuristics for evolutionary computation algorithms based on hybrid nonlinear–linear encodings [119].

Acknowledgements. We thank Douglas J. Blackiston for the tadpole picture in figure 6c and Dany Adams, Douglas Brash, Mariano Bizzarri and the members of the Levin laboratory for many fruitful discussions. This paper is dedicated to the memory of H. S. Burr, who long ago formulated a testable biophysical model of target morphology in development, regeneration and cancer suppression.

Funding statement. NIH, G. Harold and Leila Y. Mathers Charitable Foundation, US Army Medical Research and Materiel Command (USAMRMC), National Science Foundation (grant nos. GM078484, W81XWH-10-2-0098, EF-1124651, CBET-0939511).

References


