J Biosci (2023)48:14
DOI: 10.1007/s12038-023-00344-4

Perspectives

What distinguishes the elephant from *E. coli*: Causal spreading and the biological principles of metazoan complexity

PHILIP BALL
Freelance writer, London SE22, UK

(Email, p.ball@btinternet.com)

MS received 9 March 2023; accepted 28 March 2023

Keywords. Biological causation; biological complexity; gene regulation; intrinsically disordered proteins; metazoan evolution

Jacques Monod famously said that 'What is true for *E. coli* is true for the elephant.' While this might be correct in the basic sense that both use nucleic acids and proteins, it is no longer clear that they use them in quite the same way. The many qualitative differences in the biomolecular constitution and mechanisms of protozoans and metazoans, from the proportions of non-coding DNA to those of multidomain and disordered proteins and the mechanisms of gene regulation, seem to reflect different generic principles in how the two types of organism operate at the molecular and cellular levels. Here I suggest that one way to think about those differences is as a shift in the locus of biological causation – a shift that has implications for making biomedical interventions in humans.

1. Introduction

What makes life happen? This is, and always has been, arguably the biggest question for biology. Traditionally two kinds of answer have been offered. One is to delve deeply into the molecular mechanisms: to identify the genes, proteins and other molecules involved in, say, the development of a particular body part or the body's response to attack by a pathogen. Very often this approach generates complex 'wiring' diagrams of molecular interactions, in which it becomes very hard to say what controls what, which components are vital and which are peripheral, or how a molecular-scale event translates to a physiological or phenotypic one.

The alternative is to appeal to biology's sole overarching theory, Darwinian natural selection, and to explain a trait or feature in terms of its adaptive benefit. This perspective speaks to a different *kind* of explanation. It does not give an account of how the trait or feature is formed over the course of an organism's development or lifespan, but alludes to its function in making the organism viable or reproductively successful. Oddly, perhaps, biologists seem uncomfortable with any implication of *purpose* that might entail. They seek to deactivate any such terminology with scare quotes or tailor-made terms like *teleonomy*, introduced by Pittendrigh (1958) and enthusiastically adopted by Ernst Mayr, Jacques Monod, and others, and yet which arguably gains its currency only by making a straw man of teleology (Dresow and Love 2023).

http://www.ias.ac.in/jbiosci
Published online: 04 May 2023

Both approaches commonly culminate with the identification of an association between the explicandum – the trait, process, or disease – and specific genes. What part the genes actually play has often been left unclear: as philosopher of science Evelyn Fox Keller has said, such questions have typically been postponed with vague talk of 'gene action' (Keller 1995). Today the roles that genes (and other genomic elements) play in phenotypic traits are sometimes somewhat better understood, but sometimes they remain only at the level of statistical associations.

All this has left explanations of the 'how' of life couched in terms of *mechanism*, alluding more or less explicitly to a machine metaphor (Nicholson 2013). Like most metaphors, this one only becomes problematic when it becomes invisible. The suggestion that living organisms are like machines might be best read saying that, if there were indeed machines that display agency, that respond to external signals not like automata but with reference to internal states and representations, that could reconfigure their small-scale parts and interactions based on top-down high-level inputs, that could ultimately display signs of consciousness, emotion, and even wisdom – if there were truly machines like *this*, then yes indeed, life would be like those machines.

There is almost certainly no generic answer to the question of what makes life happen. If such existed, it would almost by definition have to be so vague as to be useless for posing and answering specific questions of living entities, or making predictions and testing them. Perhaps because of both the overwhelming scope and the immense difficulty of the question, biology has for the best part of a century – at least ever since Erwin Schrödinger's influential little book *What is life?* (Schrödinger 1944) – relied on a deceptively simple narrative. Life (in this view) does what it does because it possesses what Schrödinger called a 'code-script' – and what modern biology came to identify as the 'instruction manual' of the genome – that has programmed it to do what it does, albeit along pathways so baroque that it takes endless labour to elucidate them, molecule by molecule. For Ernst Mayr, the existence of a 'program' introduces 'dual causation' to the aminate realm, making life controlled not just by physical laws but by the inherited program – which, in his view, makes biology unique and autonomous while liberating it from the burden of teleology (Mayr 2004).

This was never a likely answer, however, and indeed, as Nicholson (2010) says, 'the view that genes are the primary causal agents of all the phenomena of organismic life is not well supported by the findings of contemporary biology'. For one thing, rigid bottom-up programming will be far too fragile in the face of the stochasticity of the molecular environment both inside and outside living cells. Over the past two decades or so – in the time span, (perhaps) coincidentally, since the completion of the Human Genome Project – biological research has been furnishing a better description: one that invokes organising principles at higher levels such as those of cells, tissues, and organisms.

Causation sits at the heart of this view – which speaks not only to the processes by which life unfolds but also to the question of how best to intervene when life goes awry. The deluge of genomic data gathered over the past several decades has not been translated into anything like the quantity of therapies as was hoped (Shendure *et al.* 2019), especially for diseases that are non-Mendelian (in general, that are polygenic). For monogenic diseases, many of which are rare, the gene concerned can be reasonably considered a genuine causal factor, and new techniques such as preimplantation genetic screening of embryos or (somatic cell) genome editing may be able to address the problem at its source. But when many genetic sequences are associated with a condition (and in fact most of the hits from such genome-wide association studies are in non-coding parts of the genome believed to have regulatory functions), it is far from obvious which of them, if any, point to the best therapeutic targets (such as a particular gene product). Often the answer is 'none of the above', because the associated genes or functional molecules may have highly nonspecific roles. Targeting a given gene might thus have no effect, or worse, adverse pleiotropic consequences.

Genetic data are now seen as just one small part of a much wider data set needed to extend medicine's reach while making it better tailored to the individual (Subbiah 2023). This desired data set encompasses processes at all scales between genotype and phenotype, and implicitly signifies that no scale is special in this regard – in other words, that there is no privileged causal level in living entities. But collecting data on everything is not obviously the best alternative either. That approach, originally advocated by Francis Bacon in the seventeenth century (Bacon 1878), has never proved a good way of doing science, and even with the assistance of machine learning it does not seem the most efficient or effective way to understand ourselves. Rather, we need better ways to *think* about life.

I call this situation, in which there is no privileged stratum of causation in the hierarchy of biological complexity for humans and other complex animals, *causal spreading*. I shall argue that it is likely to be a necessary outcome of the evolution of complexity in living organisms.

2. Multilevel causation and fuzzy biomolecular logic

That biological causation is not just bottom-up has long been recognized, whether implicitly or explicitly, in developmental biology. During development, a cell acquires its fate – and the corresponding epigenetic settings of its genome – via signals (chemical, mechanical, electrical) from its environment. There is no inexorable unfolding of a developmental 'code-script', but rather, a contingent developmental process that is determined collectively and provisionally and is path-dependent (Davies 2014). It is far from obvious how to formulate a causal narrative to describe this process, but it does seem clear that the genome – its activity, if not its sequence – is as much controlled by the cell and the tissue as vice versa (Martinez Arias 2023).

Here is an example of this interplay of top-down and bottom-up control. The formation of a villus in the epithelial tissue of the gut is induced by the release of the protein Sonic hedgehog (SHH). Above a certain threshold concentration, SHH triggers a signalling pathway that causes epithelial cells to undergo an epithelial–mesenchymal transition (EMT), reducing cell–cell adhesion and the bending modulus of the tissue layer so that it buckles and forms the nascent villus protrusion (Shyer *et al.* 2015). One might be tempted to say that SHH *causes* villus formation.

Or, should we say instead that the EMT is what really produces a villus, and that SHH is just one possible causes of this change in cell properties? After all, buckling of the embryonic ectoderm in the formation of the neural tube and crest is a comparable process, also reliant on the loosening of cell contacts via the EMT – but it is in that case triggered by other morphogens including Wnt and BMP.

There is, in any event, nothing about SHH signalling or the EMT that encodes the shape of a villus. Rather, the form emerges from a feedback process: buckling confines the SHH protein to the region within the protrusion and thus raises the local concentration above the threshold needed to trigger the EMT (figure 1). One might equally say that villus formation is its own cause: it is a self-amplifying process akin to that the Mullins–Sekerka instability that produces dendritic fingers in crystal growth (Langer 1980).

So which is it? It makes little sense to promote one 'cause' over the other. Causation is *spread* in this process from molecules (SHH) to cell states (the EMT) to tissue mechanics (buckling).

On the one hand, there is 'nothing to see here': this is normal, uncontroversial developmental biology. But I believe a deeper framing of this kind of process can provide a better and, importantly, more useful way to understand how life works.

The spreading of causation seems to be a central feature of the molecular mechanisms that govern signalling and gene regulation in metazoan cells. Changes in gene regulation can be triggered by signals coming from outside the cell's nucleus, and indeed outside the cell itself: the arrival of a hormone in the bloodstream, say (which might induce a cascade of kinase activity), or developmental signals from surrounding tissues. These messages are conveyed through networks of interacting proteins in the cytoplasm. Likewise, protein networks interpret and

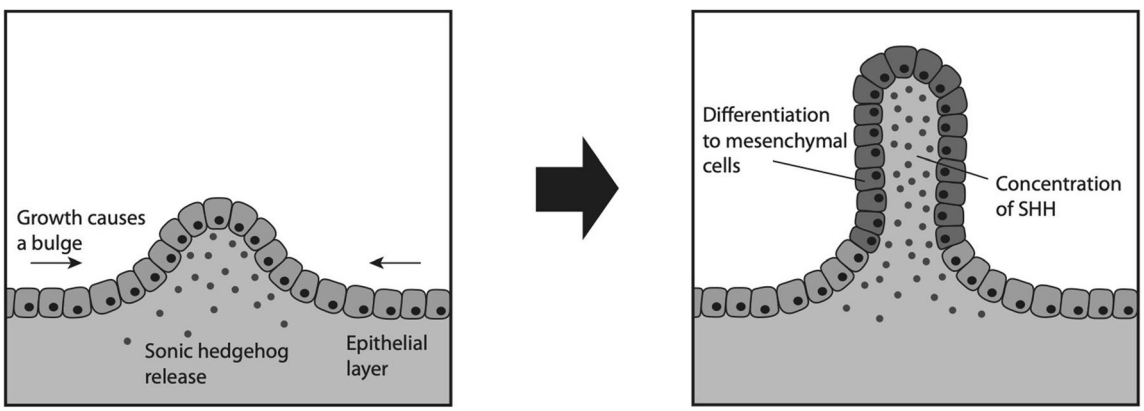

Figure 1. Formation of a villus by localized concentration of the signalling protein Sonic hedgehog (SHH).

respond to the messages coming in the other direction: from the nucleus, in the form of altered expression levels of genes. The signalling goes both ways.

Furthermore, the signalling within protein interaction networks appears, at least in some cases (often involving key developmental gene products), not to have the specificity of connectivity or interaction – molecule A speaks solely and uniquely to molecule B – that old analogies between biology and computing and information theory asserted. Instead, the informational principles involved seem uniquely biological, and attuned specifically to spread responsibility for the outcome so that the burden is not, as it were, shouldered uniquely by any molecule that is hostage to the stochasticity of the cell environment.

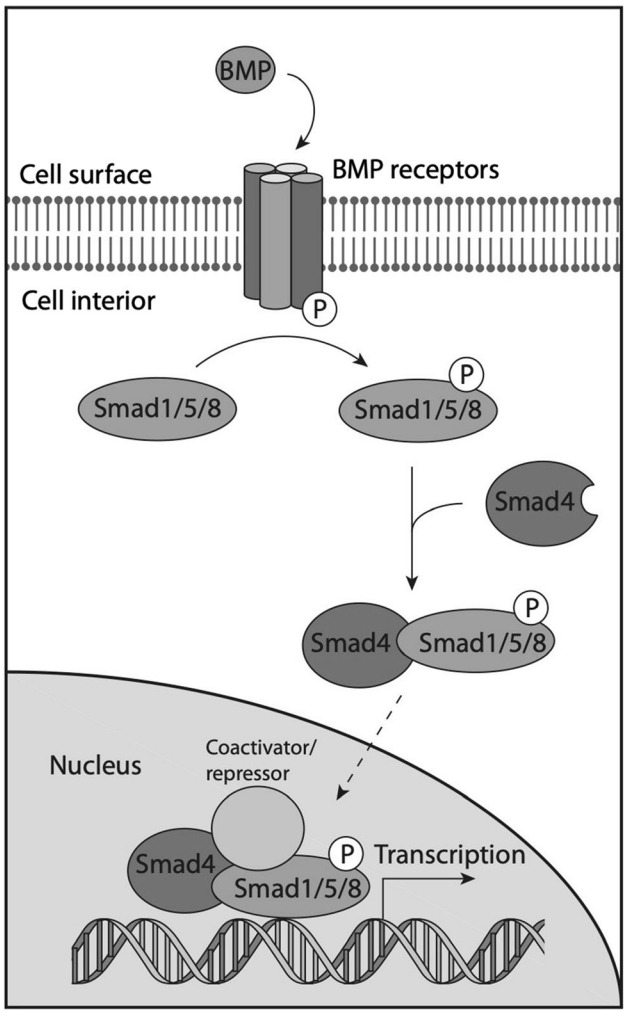

Figure 2. How BMP proteins take part in cell signalling. The binding of a BMP to its receptor protein in a cell membrane triggers phosphorylation of a Smad protein on the other side of the membrane. This modified Smad, in combination with other varieties of Smads, can then regulate the expression of a gene in the nucleus.

Consider the BMP signalling pathway, which is implicated in a wide variety of mammalian developmental processes: in gastrulation and the formation of bone, cartilage, kidney, eye, and early brain tissue, wound repair and the reshaping of blood vessel networks (Wang *et al.* 2014). BMPs are key hub molecules in several protein networks through which signals are routed that direct cell behaviour and fate. The pathways mediated by BMPs begin at the surface of a cell, where these proteins bind to receptor proteins spanning the membrane, changing the receptor shape allosterically in a way that registers the binding event at the other end of the receptor, on the inside surface of the membrane. This change is in turn registered by Smad proteins, which convey the signal to the nucleus and act as transcription factors (figure 2).

Different BMPs – there are eleven or so for mammals, each encoded by a different gene – can bind to the receptors and convey different signals. BMP proteins are dimeric, while their receptors typically have four subunits. The BMP dimers do not each have a designated receptor to which it binds like lock and key; rather, each BMP might stick to several different combinations of receptor subunits, with varying degrees of avidity. It is a combinatorial system (figure 3).

With this promiscuity of ligand–receptor interactions, how can the BMP pathway deliver a specific message to guide a cell's fate? Although they are not highly selective, the various BMPs do have preferences for certain receptors, and Elowitz and coworkers (Antebi *et al.* 2017; Klumpe *et al.* 2022) have mapped

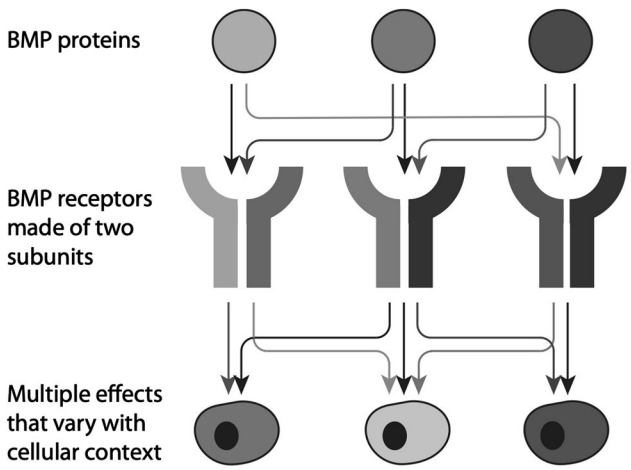

Figure 3. In the combinatorial BMP signalling pathway, different BMPs can dock into different receptors, and different combinations will have various effects on the cells. Here, for simplicity, each BMP is shown as a single entity, and the receptors as pairs of subunits; in reality each of those entities is itself dimeric.

out the binding propensities between the major mammalian (mouse) BMPs and their receptor subunits. These preferences were inferred from the strength of the activating effect on Smad proteins inside the cells, as revealed by coupling Smad activity to GFP expression (Klumpe *et al.* 2022).

The interactions are promiscuous, but not arbitrary. Some BMPs have much the same effect as one another and are interchangeable, while others do not. Sometimes two different components might have independent effects so that their combined effect is a simple sum. In other cases the effects might mutually amplify or inhibit each other. In general the BMPs can be grouped according to whether they will have the same effect in a given context, such as a certain cell type. A pair of BMPs might substitute for each other in one type of cell but not in another. This is consistent with how the biological effect of a given BMP differs in different tissues. For example, the protein denoted BMP9 can substitute for BMP10 in the BMP pathway that leads to formation of blood-vessel networks, but not in the pathway involved in heart development (Chen *et al.* 2013). Such context-dependence is hard to reconcile within a paradigm that supposes molecular communication to depend on highly selective recognition and binding.

Why does BMP signalling work in this apparently complicated way? The answer may be that it offers more for less. With only one way a given molecular component can work, there are only a few combinations of the elements that will be functional. Within a combinatorial system there are many more functional combinations. One potential advantage of such combinatorial diversity of outputs is that it allows the molecules to convey distinct messages to a wider variety of cell types – and so, in a developing tissue, to produce more complex patterning with a small repertoire of signalling molecules (Su *et al.* 2022). It is, in other words, a principle well suited to an organism with diverse tissue types.

The combinatorial and promiscuous signalling used in the BMP pathway might represent a widespread design principle of the molecular wiring of cells. There is a loose parallel with the olfactory system. For humans there are around 400 receptor proteins lining the epithelial membranes of the nose, which together can discriminate between a vast number of odors (one estimate puts it at one trillion) (Sharma *et al.* 2019). That range would not be possible if each odorant molecule had to be uniquely recognized by its own dedicated receptor. Instead, the receptors seem to bind odorants somewhat promiscuously with different affinities, and the output signal sent to the brain's smell center is then determined by combinatorial rules. Perhaps the most useful analogies for how cells work are themselves biological, such as are found in olfaction, cognition (Shapiro 2021), or the immune response.

3. Looseness of eukaryotic gene regulation

The same kind of fuzzy combinatorial logic operates in eukaryotic gene regulation. Mammalian chromatin is divided into distinct compartments, each with a distinct network of contacts between different parts of the chromatin, as well as different classes of epigenetic marks (Liebermann-Aiden *et al.* 2009; Rao *et al.* 2014). Each compartment contains many dense clusters of DNA, proteins and RNAs called topologically associating domains (TADs) (Beagan and Phillips-Cremins 2020; Tena and Santos-Pereira 2021). These clusters appear to mediate gene regulation (Agelopoulos *et al.* 2021), and typically contain enhancer sequences of DNA on chromatin loops, which influence gene expression (such enhancers are often remote from the genes they regulate, and so are brought into proximity with them by looping), along with transcription factors and other proteins with low binding affinity (including RNA-binding proteins), regulatory RNA molecules, and cofactors (figure 4). Epigenetic markings on chromatin, particularly histone methylation, are important for the formation of these regulatory

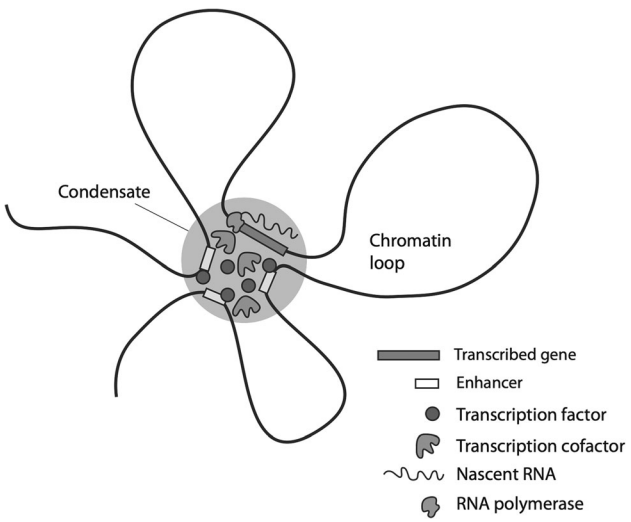

Figure 4. Schematic illustration of a topologically associating domain of chromatin, thought to be a liquid-like 'condensate' that constitutes a distinct phase from the surrounding fluid.

hubs. There is, in other words, a rich association of diverse signals involved in the regulatory process.

In contrast to the classic regulatory system of the bacterial *lac* operon (Cunliffe 2005), eukaryotic regulatory hubs are thus remarkably heterogeneous, non-specific, and dynamic. Tjian and coworkers (Chen *et al.* 2014) have found that components within a TAD stay bound to one another for just a few seconds. How can a low-concentration protein ever get together with all its partners to trigger expression of a gene, when everything is moving at such a rapid pace? One popular view is that, rather than forming a precisely structured assembly, the components of a TAD gather into a liquid-like droplet enveloping the gene being regulated, a process called liquid–liquid phase separation (Hyman *et al.* 2014; Shin and Brangwynne 2017). Within these 'condensates', the proteins can repeatedly bind to and unbind from one another and from DNA while remaining co-localized (Palacio and Taatjes 2022). Regulation is then a cooperative affair involving many repeated binding events. Such a collective process is consistent with the fact that many transcription factors are non-specific in their binding to DNA: they will stick more or less strongly to various different sequences. The ENCODE project (ENCODE *et al.* 2020) identified more than 630,000 genomic regions that look like potential binding sites (*cis* regulatory elements), comprising about 8% of the entire human genome.

This picture of gene regulation via the formation of liquid droplets might explain how a protein that has a very low concentration in the cell as a whole can develop a locally high concentration at the gene it regulates. (Prokaryotic regulation, as in the classic *lac* operon, faces the same problem but seems to have a different solution in that case: nonspecific binding of a repressor to DNA followed by one-dimensional diffusion along the genome (Von Hippel *et al.* 1974).) Yet compelling evidence for droplets forming *in vivo* amidst the tangled mass of chromatin in the cell nucleus is difficult to find. Rather than being persistent droplets, condensates might be more transient and dynamic 'hubs' (McSwiggen *et al.* 2019). The key point, however, is that these are loose and rather non-specific associations, not precisely defined complexes that must be assembled completely and robustly before they can function. Information is transmitted collectively (in ways not yet fully understood) rather than by specific molecular-recognition processes. Thus, gene activity emerges as a consensus process that can integrate a variety of inputs, some of which might be contradictory – activating or inhibiting transcriptional signals, say. Again, the computational logic underlying life is fuzzy and stochastic, and seemingly analog in character.

4. Role of protein disorder

Many of the proteins that act as hubs in the interaction networks for both cell signalling and gene regulation have a high degree of intrinsic disorder (Wright and Dyson 2015; Kulkarni and Uversky 2018). Disordered proteins have a propensity to form many indiscriminate and transient interactions with other molecules, and so seem to be ideally suited for promoting the formation of condensates (Choi *et al.* 2020) and for creating promiscuity and combinatorial logic in signalling pathways: for producing the 'weak linkages' in interaction networks that Kirschner and Gerhart (2005) identify as central to many regulatory mechanisms.

The proportion of disordered segments in the entire human proteome may be as high as 37–67% (Deiana *et al.* 2019; Uversky 2019). Disorder seems particularly prevalent in many of the most important proteins in the molecular ecology of the cell, and seems to be a characteristic of metazoan proteomes in particular: intrinsically disordered proteins are significantly less prevalent (perhaps as little as 4% or so) in many bacterial proteomes. Disorder thus seems *functional* in the evolution of metazoa. Pappu and coworkers have identified four distinct classes of disorder, ranging from chains that are more or less open and formless to structures that are relatively compact, like loose blobs (Das *et al.* 2015). These differences are reflected in the protein sequence and are related to function: for example, compact globules seem to bind effectively to DNA, the proximity of the different parts of the chain helping to create a collectively strong interaction.

Disorder is also prominent in RNA-binding proteins (RNABPs), which play an important role in the splicing of mRNA – the removal of introns – after it has been transcribed, as well as in a variety of other cell processes (Lee and Rio 2015; Gebauer *et al.* 2021). Alternative splicing – in effect, which protein is produced from a gene – is controlled by a loose network of interactions between regulatory RNAs and low-affinity RNABPs (Lee and Rio 2015). The rules of this process are not well understood, but the low specificity of the interactions may help to make it highly responsive to the overall state of the cell, so that for example different proteins are made from the same gene in different cell and tissue types. The allows the 'output' of the genome to be highly and rapidly responsive and

adaptable to changes in its environment and in the state of the cell – a sensitivity to 'top-down' information that is likely to be vital to the robustness of metazoan development. Again, the key to that responsiveness is not rigid and highly specific channels of information flow (favoured by highly selective molecular interactions) but an ability to *integrate* many diverse signals into an optimal and context-dependent output.

Structural disorder is a key enabler of protein function. Hilser and Thompson (2007) propose that disorder is conducive to allosteric effects, because it does not require precise engineering; rather, binding of a ligand produces subtle shifts in the many different conformations to which a disordered protein has access, and these collectively 'spread the word'. In this way, disordered proteins can become versatile connectors between different chains of interaction that convey signals in the cell, making them excellent hubs in the networks of molecular interactions involved in signal transduction and regulation. Even though – indeed, precisely because – intrinsically disordered proteins interact only weakly and transiently with many other molecules, they enable cells to respond quickly to a change in circumstances, giving access to a wide variety of possible routes for transmitting and directing signals that are not pre-programmed into the system. Disordered proteins are also versatile communicators in signalling pathways because their loose shape can be readily altered by phosphorylation (Wright and Dyson 2015).

The disorder and promiscuous ligand binding of many human proteins does not just mean that the molecular mechanisms of our cells are somewhat messier and looser than those typically seen in prokaryotes; it implies key differences in the structures of informational networks and principles. Gene sequences do not program function into disordered proteins by specifying a particular molecular structure and shape in a precise, machine-like way. Rather, what a disordered protein *means for the cell* is mutable with the state of the whole cell – it becomes responsive to top-down information. In this way the information ecosystem within which proteins operate is not that of a *closed* genetic blueprint but is *open*. It makes no sense to suppose that any given level of this complex system is more 'in control' than any other.

These principles are not confined to eukaryotes, but seem to be more widely deployed in them. The aforementioned *lac* repressor protein of *E. coli*, for example, is also a partly disordered, multidomain protein that binds with low specificity and relies on allostery; only when the dimeric protein binds to the target sequence does the protein become ordered (Seckfort *et al.* 2020). Yet it seems possible that making promiscuous, reconfigurable networks does not merely convey advantages in more complex organisms but perhaps is the only way multicellular eukaryotes *can* work, if the system is to be robust against ineluctable randomness and unpredictability in its fine details. Cellular systems are noisy: molecular encounters in this crowded, jostling environment are very much a matter of chance, and there are also random fluctuations in the amount of different proteins that get produced from moment to moment (Losick and Desplan 2008). No two cells are ever in wholly identical transcriptional states at any given instant, even when they both ostensibly have the same role – as muscle or kidney cells, say (Wagner *et al.* 2018).

A cell in which each component is wired specifically to another would be highly vulnerable to such uncontrollable variability. If successful functioning depends on ultra-precise duplication of the network wiring, any mutations are likely to be deleterious. Combinatorial logic, on the other hand, has a certain amount of sloppiness that can absorb (and even exploit) such variation. It makes biological systems robust.

5. Implications for evolution

In the Neo-Darwinian Modern Synthesis, inheritance is separated from development – a separation asserted by the Weissman barrier between somatic and germ cells, and inherent in Crick's Central Dogma of biological information flow (Crick 1970), which, according to Walsh, 'underscored the distinction between inheritance and development' (Walsh 2015, p. 79). But the neglect of development in modern evolutionary theory never went unchallenged (Waddington 1961), and more recently the calls for recognizing morphology as an active influence on evolution rather than a consequence of it have been more insistent (Newman 1992; Kirschner and Gerhart 2005; Walsh 2015). They argue that evolutionary dynamics simply cannot be understood by looking only at statistical changes in allele frequencies in populations; it demands a consideration of the generic morphological processes for which genes supply the molecular resources.

To follow just one thread: Kirschner and Gerhart (2005) argue that phenotypes that are relatively robust to mutation via canalization of development – and that are thus evolvable – may be selected for such evolvability demands more than random genetic mutation and natural selection of the resulting phenotypes

(Wagner 2015). To be viable at all without undermining development, a genetic mutation needs to be insulated from the phenotype by organizational layers that can integrate it into a coherent whole. Consider Darwin's Galapagos finches, their beaks so seemingly well adapted to the specific function they had to fulfil in different evolutionary niches. At face value it is not easy to evolve a beak shape and keep it functional: how does the lower beak, say, not becoming outsized with respect to the upper one? How are small, gradual changes to the beak kept proportionate to independent changes in the head and musculature? Developmental mechanisms smooth out such potential inconsistencies: a single signalling molecule (in the case of avian beaks, BMP) influences the size of the whole beak. The buffering provided by the higher levels of organization reduces the likely lethality of genetic change (Gerhart and Kirschner 2007; Kirschner 2013). At the same time, those higher levels provide ways in which small genetic changes – François Jacob's 'tinkering' (Jacob 1977) – in regulatory pathways can elicit significant variation in phenotypes, rather than correspondingly tiny alternations of form. None of this in itself conflicts with Darwinian evolution, but it does suggest that the Modern Synthesis might need expanding to accommodate such a developmental perspective (Walsh 2015).

Gerhart and Kirschner (2007) suggest that such variation involves the tweaking of a 'core system' within genomes that supplies the basic ingredients for anatomical development of all higher animals: a toolkit that doesn't specify a particular body shape, but rather, enables cells to develop into coherent and integrated systems of tissues. This anatomy-generating core system is highly conserved in metazoa. The result will always be a coherent limb ending in digits, say, made from vascularized tissues and strengthened by a skeleton – but it could be an arm, a fin, a wing. In other words, the core system remains a 'body-generating' network in the face of mutation, while being capable of generating diverse yet coherent phenotypes. Development might then be regarded in terms of dynamical attractors (Sáez *et al.* 2022a, b): generally rather robust to mutational 'noise', but characterized by occasional, abrupt shifts to a wholly different morphological basin.

Crucially, these core processes derive their robustness largely from the *weakness* of their regulatory linkages. They can accommodate new patterns of regulation by virtue of the low specificity with which the molecules interact with one another in networks, giving reliable and coherent developmental outputs that are insensitive to the fine details of the molecular discourse. Promiscuous, highly interconnected protein networks also promote the ability of the organism to acquire useful new capacities by evolution. If promiscuous binding allows one protein to substitute for another, the network can develop new functions without losing old ones. It seems likely that metazoans have *evolved this evolvability.* Payne and Wagner have shown that the promiscuous binding of transcription factors can indeed promote both robustness to mutations and the ability to evolve (Payne and Wagner 2014). The selective pressure on evolvability – derived from promiscuity in the molecular pathways of the cell mechanisms – might be expected to be far more important for large animals than for bacteria, because they have much smaller populations and will not so easily weather an environmental change via some fortuitous mutation arising rapidly in a vast number of offspring.

6. Causal spreading

Understanding causation in complex systems like this is challenging. In general, efforts to quantify causation consider to what extent a putative causal factor is both sufficient (the outcome becomes obligatory in the presence of the factor) and necessary (the outcome will not happen in its absence). Some such metrics of causality show that, in complex hierarchical systems, it may be localized more in higher levels of the system rather than flowing from the bottom up (Hoel *et al.* 2013): there may be a (quantifiable) degree of *causal emergence*.

Comolatti and Hoel (2022) have shown that more than a dozen such measures of causation *all* reveal causal emergence in some complex systems: it seems to be a real property of such systems and not a quirk of the specific causal metric. Such causal emergence instils noise reduction: independence of the outcome on random fluctuations or chance events at the microscopic level. It therefore makes perfect sense that evolution would 'design' complex organisms to display causal emergence. It amounts to fitting the scale of the causes to that of the effects. Morphologies and other phenotypes, in this view, may be influenced by specific genetic changes but are not acutely sensitive to them. This is just what is seen in many gene knock-out studies (Morange 2001). Traditionally such robustness has often been attributed (without real evidence) to redundance of molecular pathways. But increasingly it seems more likely to be an emergent property of

regulatory and signalling networks, cell–cell interactions, and other higher-level phenomena.

This is not to say that all causation in complex organisms is emergent. Evidently, single-gene mutations do sometimes have a significant impact on a phenotype, and a drug that hits a specific protein can sometimes disrupt a physiological response. Evolution does not so much shift all causation to higher levels as spread it among the various levels. We might anticipate that there are reasons why more or less emergence is best for a given process – but if so, we do not know in general what they are.

Hoel *et al.* (2020) have investigated how and where causal emergence arises in the networks of protein interactions (the interactomes) of a wide range of organisms. They looked at around 1,500 species of bacteria, 11 archaea, and 190 eukaryotic species. Using a measure called 'effective information' to quantify causation (Hoel *et al.* 2013), they searched for informative macroscales in the networks – situations where a group or cluster of protein–protein interactions could be replaced by a single macro-node in the network that does the same job as the collective. Such macro-nodes can be considered autonomous units in producing the observed outcome at the level of the phenotype: they will function reliably even if there is some variability or noise among their component parts. They become genuine causal entities that operate at a higher level of organization.

The analysis reveals significantly more causal emergence – more informative macroscales – for eukaryotes than for prokaryotes (figure 5). In this regard – and also, as we have seen, in the mechanisms underpinning it – what is true for *E. coli* appears not to be the case for elephants. (That putative equivalence, while commonly attributed to Monod in the 1950s, in fact seems to stem from the Dutch microbiologist Albert Jan Kluyver three decades earlier (Kluyver and Donker 1926; Friedmann 2004).)

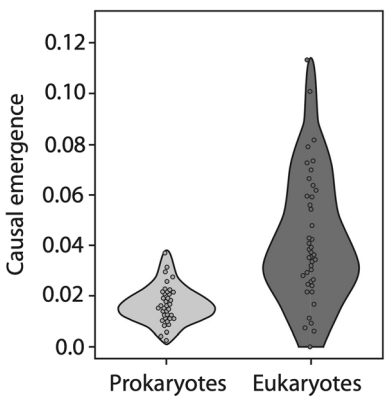

Figure 5. Causal emergence is more evident in eukaryotes, especially multicellular complex organisms, than in prokaryotes. In other words, evolution produces 'causal spreading', in which causation is increasingly conferred on higher organization levels. Adapted from Hoel et *al.* (2020).

Did this shift from primarily microscale to higher-level causation in interactome networks result from the switch to multicellularity, or did it enable that? Some clues might be gleaned from species at the borderline of unicellularity and multicellularity: for example, the single-celled eukaryote *Capsaspora owczarzaki*, one of the closest evolutionary relatives of the first multicellular animals, such as sponges, cnidarians and ctenophores. *Capsaspora* has more genes involved in regulatory functions (generally by encoding transcription factors) than any other single-celled organism. The networks that these transcriptions factors govern are often found too in animals: the networks were already primed and ready to go before true multicellularity emerged (Sebé-Pedrós *et al.* 2016). The switch to multicellularity seems to involve the appearance not of more primary genetic resources – more or different genes – but of new ways to regulate them.

In effect, then, the more complex multicellular organisms that appear later in evolutionary history tend to assign causal roles to higher levels of organization in their networks. In this way, these organisms can tolerate more noise and indeterminism in the microscales, because those scales are not the primary determinant of phenotypic outcomes such as body shapes and behaviours.

I call this phenomenon *causal spreading* – a concept first introduced in the context of the philosophy of mind (Clark and Chalmers 1998). It means that if we want to understand the mechanisms behind some key evolutionary shifts – for example, the emergence of complex body shapes and lifestyles in the Cambrian explosion, the emergence of nervous systems and of new modes of cognition, and the divergence of mammals and other vertebrates – genomes are the wrong place to look. All we will find there are echoes of the true causal factors that happened at a higher level of organization – in particular, changes within the networks of interaction between and regulation of the molecular components.

This interpretation of the evolution of organismal complexity is supported by the analysis by Lowe *et al.* of changes in genome sequences called nonexonic elements (sequences that fall outside of exons) in humans, cows, mice, and two types of fish (sticklebacks and Japanese medaka) (Lowe *et al.* 2011). Nonexonic elements (NEEs) that are highly conserved

are likely to have a functional (probably regulatory) role – although it has been difficult to discern what the role of such 'ultraconserved' genomic elements might be (Kasbekar 2023; Sabarinadh *et al.* 2003). The phylogenetic comparison of NEEs revealed three distinct eras of change over the past 650 million years. Until about 300 million years ago, when mammals split from birds and reptiles, changes in regulation seem to have happened mostly in parts of the genome close to transcription factors and the key developmental genes that they control. Then between 300 and 100 million years ago those changes diminished, and instead there were changes near genes that code for cell surface receptors: a shift in cell–cell communication. Finally, since 100 million years ago, as placental mammals developed, the regulatory changes seem to be associated with mechanisms for modifying protein structure after translation, especially for proteins that are associated with signal transduction within cells. These changes during the evolution of complex multicellular organisms can be regarded as shifts in the *locus of causation* of phenotypes.

7. Conclusions

It is all too easy in biological research to end up seeing 'nothing but trees'. When that seems to be the case, the impulse is all too often to seek more trees. If, for example, promiscuous interactions make it hard to causally connect the outcomes of a cell process to the molecular components seemingly involved in it, the understandable impulse is to suppose that more data are needed. The same applies if outcomes are found to be context-dependent – if, say, the effect of a signalling pathway depends on the state of the cell in which it happens, or if a phenotype is found to be extremely polygenic. Analogous complexity is typical in the social sciences, where it might be naively supposed that differences in outcomes demand a detailed characterization of the temperamental differences in the individuals involved in them. But in the presence of causal spreading, such delving into details does not necessarily bring us any closer to a causal explanation or model.

Causal spreading seems to appear in complex organisms in parallel with canalization (Waddington 1942) – the tendency for outcomes (developmental or physiological) to be broadly prescribed and insensitive to details and fluctuations. Despite the wide diversity of transcriptional profiles in cells within a given tissue, for instance, they remain recognizably hepatocytes, astrocytes, fibroblasts and so on. This seems to be the result of dynamical landscapes of gene expression that have just a few wide basins of attraction (Sáez *et al.* 2022a, b), typically dictated by the activities of just a few key genes (Huang *et al.* 2007). The two processes probably go hand in hand as mechanisms for attaining robustness of outcomes in the face of molecular stochasticity. In other words, the problem confronted by Schrödinger of how biological predictability ('order') arises from molecular disorder is solved not by the bottom-up imposition of order in a 'code-script' but by mobilizing collective phenomena leveraged by general principles of self-organization and complex dynamics, enacted on high-dimensional dynamical landscapes that are amenable to dimensional reduction to create canalization. This looks like a general prescription for robustness and evolvability; but in metazoa it becomes a necessity, because of the need for cell- and genome-scale events to be responsive to those at the level of tissues, organs, and whole organisms.

Causation and canalization can occur at many levels of the biological hierarchy. The implications are not just scientific or medical but even philosophical and moral. The 'genetic blueprint' picture implies the notion of a plan and of errors that cause deviations from it. This is not wrong, exactly: it does not seem unreasonable to speak of sickle-cell disease in terms of a malfunctioning HBB gene, for example. But in general, what biological processes produce is not a fixed target but a palette of possibilities from which the selection is contextual, contingent and mutable. Too normative a picture of biology at the organismal level is prejudicial. Is dwarfism a developmental 'error', or simply a possible outcome of the developmental process – an outcome to which our society creates barriers? Are autistic-spectrum conditions (which, it might be argued, seem too widespread to be simply 'aberrant' phenotypes) also 'errors'? Where do we draw the lines between 'ideal' targets and deviations from them? Is any of us truly representative of an 'ideal' developmental outcome, or for that matter of an 'ideal' genome?

A recognition of the casual spreading of human biology can lend greater sophistication to such discussions of societal norms and categories. The common assertion that 'sex is biological' all too often implies that it can be ascribed a single biological determinant: a locus of binary division. But there is none – not at the genomic, chromosomal, anatomical, hormonal or even neurological level. Sex is another phenotypic trait that becomes canalized, but there is no level of the biological or developmental landscape that uniquely defines the destination. Our societal discussion would be richer for acknowledging it.

I want finally to stress how important it is that these considerations be operationalized in ways that are useful and productive for molecular, cell, developmental and evolutionary biology. It is no surprise that many biologists respond with a shrug to abstract discussions about causation or dynamical landscapes, since these perspectives rarely had much to offer for benchtop experiments. But that can and will (and must) change. For example, recent experiments show that a dynamical-landscape description of cell-fate determination can be quantified and constrained by real-world data – perhaps in a way that identifies which are the effectual knobs and dials that can program cell states (Sáez *et al.* 2022a, b). And efforts to extend synthetic biology to eukaryotes (Shao *et al.* 2018; Zhu *et al.* 2022) and to create 'multicellular engineered systems' (Kamm *et al.* 2018) from 'agential' living materials (Davies and Levin 2023) will surely demand a better understanding of what factors control morphology and functionality at levels beyond the genomic. As Richard Feynman famously said, 'What I cannot create, I do not understand'. The maxim can also be inverted as a reminder that systematic realization of goals – in biomedicine, tissue engineering, biotechnology – will always remain difficult while the principles are not understood.

Author contribution

P.B. is the sole author of the paper.

Funding

I have no sources of funding for this work.

Declarations

Conflict of interest I have no conflicts of interest to declare.

References

Agelopoulos M, Foutadakis S and Thanos D 2021 The causes and consequences of spatial organization of the genome in regulation of gene expression. *Front. Immunol.* **12** 682397

Antebi YW, Linton JM, Klumpe H, *et al*. 2017 Combinatorial signal perception in the BMP pathway. *Cell* **170** 1184–1196

Bacon F 1878 *Novum organum* (Ed.) T Fowler (Oxford: Clarendon Press)

Beagan JA and Phillips-Cremins JE 2020 On the existence and functionality of topologically associating domains. *Nat. Genet.* **52** 8–16

Chen H, Brady Ridgway J, Sai T, *et al*. 2013 Context-dependent signaling defines roles of BMP9 and BMP10 in embryonic and postnatal development. *Proc. Natl. Acad. Sci. USA* **110** 11887–11892

Chen J, Zhang Z, Li L, *et al*. 2014 Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156** 1274–1285

Choi J-M, Holehouse AS and Pappu RV 2020 Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49** 107–133

Clark A and Chalmers D 1998 The extended mind. *Analysis* **58** 7–19

Comolatti R and Hoel E 2022 Causal emergence is widespread across measures of causation. *arXiv http://www.arxiv.org/abs/2202.01854*

Crick FHC 1970 Central dogma of molecular biology. *Nature* **227** 561–563

Cunliffe L 2005 Maximum control. *Nat. Rev. Mol. Cell Biol.* **6** S8

Das RK, Ruff KM and Pappu RV 2015 Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32** 102–112

Davies JA 2014 *Life unfolding* (Oxford: Oxford University Press)

Davies J and Levin M 2023 Synthetic morphology with agential materials. *Nat. Rev. Bioeng.* **1** 46–59

Deiana A, Forcelloni S, Porrello A, *et al*. 2019 Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One* **14** e0217889

Dresow M and Love AC 2023 Teleonomy: revisiting a proposed conceptual replacement for teleology. *Biol. Theor. https://doi.org/10.1007/s13752-022-00424-y*

ENCODE Project Consortium *et al.* 2020 Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583** 699–710

Friedmann HC 2004 From 'butyribacterium' to '*E. coli*': an essay on unity in biochemistry. *Perspect. Biol. Med.* **47** 47–66

Gebauer F, Schwarzl T, Valcárel J, *et al*. 2021 RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* **22** 185–198

Gerhart J and Kirschner M 2007 The theory of facilitated variation. *Proc. Natl. Acad. Sci. USA* **104** 8582–8589

Hilser VJ and Thompson EB 2007 Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. USA* **104** 8311–8315

Hoel EP, Albantakis L and Tononi G 2013 Quantifying causal emergence shows that macro can beat micro. *Proc. Natl. Acad. Sci. USA* **110** 19790–19795

Hoel E, Klein B, Swain A, *et al.* 2020 Evolution leads to emergence: an analysis of protein interactomes across the

tree of life. *BioRxiv* https://doi.org/10.1101/2020.05.03.074419

Huang S, Guo Y-P, May G, *et al*. 2007 Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305** 695–713

Hyman AA, Weber CA and Jülicher F 2014 Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30** 39–58

Jacob F 1977 Evolution and tinkering. *Science* **196** 1161–1166

Kamm RD, Bashir R, Arora N, *et al*. 2018 The promise of multi-cellular engineered living systems. *APL Bioeng.* **2** 040901

Kasbekar DS 2023 Mysteries in our genome. *J. Genet.* **102** 1

Keller EF 1995 *Refiguring life* (New York: Columbia University Press)

Kirschner M 2013 Beyond Darwin: evolvability and the generation of novelty. *BMC Biol.* **11** 110

Kirschner M and Gerhart J 2005 *The plausibility of life: Resolving Darwin's dilemma* (New Haven & London: Yale University Press)

Klumpe HE, Langley MA, Linton JM, *et al*. 2022 The context-dependent, combinatorial logic of BMP signaling. *Cell Syst.* **13** 388–407

Kluyver AJ and Donker HJL 1926 Die Einheit in der Biochemie. *Chem. Zelle Gewebe.* **13** 134–190

Kulkarni P and Uversky VN 2018 Intrinsically disordered proteins: the dark horse of the dark proteome. *Proteomics* **18** e1800061

Langer JS 1980 Instabilities and pattern formation in crystal growth. *Rev. Mod. Phys.* **52** 1

Lee Y and Rio DC 2015 Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* **84** 291–323

Liebermann-Aiden E, van Berkum NL, Williams L, *et al*. 2009 Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326** 289–293

Losick R and Desplan C 2008 Stochasticity and cell fate. *Science* **320** 65–68

Lowe CB, Kellis M, Siepel A, *et al*. 2011 Three periods of regulatory innovation during vertebrate evolution. *Science* **333** 1019–1024

Martinez Arias A 2023 *The master builder* (New York: Basic Books)

Mayr E 2004 *What makes biology unique?* (Cambridge: Cambridge University Press)

McSwiggen DT, Mir M, Darzacq X, *et al*. 2019 Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev.* **33** 1619–1634

Morange M 2001 *The misunderstood gene* (Cambridge, MA: Harvard University Press)

Newman SA 1992 Generic physical mechanisms of morphogenesis and pattern formation as determinants in the evolution of multicellular organization. *J. Biosci.* **17** 193–215

Nicholson DJ 2010 Biological atomism and cell theory. *Stud. Hist. Phil. Biol. Biomed. Sci.* **41** 202–211

Nicholson DJ 2013 Organisms ≠ machines. *Stud. Hist. Phil. Biol. Biomed. Sci.* **44** 669–678

Palacio M and Taatjes DJ 2022 Merging established mechanisms with new insights: Condensates, hubs, and the regulation of RNA polymerase II transcription. *J. Mol. Biol.* **434** 167216

Payne JL and Wagner A 2014 The robustness and evolvability of transcription factor binding sites. *Science* **343** 875–877

Pittendrigh CS 1958 Adaptation, natural selection, and behaviour; in *Behavior and evolution* (Eds.) A Roe and GG Simpson (New Haven: Yale University Press)

Rao SSP, Huntley MH, Durand NC, *et al*. 2014 A three-dimensional map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159** 1665–1680

Sabarinadh C, Subramanian S and Mishra RK 2003 Extreme conservation of non-repetitive non-coding regions near *HoxD* complexes of vertebrates. *Genome Biol.* **4** P2

Sáez M, Briscoe J and Rand DA 2022a Dynamical landscapes of cell fate decisions. *J. R. Soc. Interface* **12** 20220002

Sáez M, Blassberg R, Camacho-Aguilar E, *et al*. 2022b Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. *Cell Syst.* **12** 12–28

Schrödinger E 1944 *What is life?* (Cambridge: Cambridge University Press)

Sebé-Pedrós A, Ballaré C, Parra-Acero H, *et al*. 2016 The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell* **165** 1224–1237

Seckfort D, Lynch GC and Pettitt BM 2020 The lac repressor hinge helix in context: the effect of the DNA binding domain and symmetry. *Biochim. Biophys. Acta Gen. Subj.* **1864** 129538

Shao Y, Lu N, Wu Z, *et al*. 2018 Creating a functional single-chromosomal yeast. *Nature* **560** 331–335

Shapiro JA 2021 All living cells are cognitive. *Biochem. Biophys. Commun.* **564** 134–149

Sharma A, Kumar R, Aier I, *et al*. 2019 Sense of smell: structural, functional, mechanistic advancements and challenges in human olfactory research. *Curr. Neuropharmacol.* **17** 891–911

Shendure J, Findlay GM and Snyder MW 2019 Genomic medicine – progress, pitfalls, and promise. *Cell* **177** 45–57

Shin Y and Brangwynne CP 2017 Liquid phase condensation in cell physiology and disease. *Science* **357** eaaf4382

Shyer AE, Huycke TR, Lee C, *et al*. 2015 Bending gradients: how the intestinal stem cell gets its home. *Cell* **161** 569–580

Su CJ, Murugan A, Linton JM, *et al*. 2022 Ligand-receptor promiscuity enables cellular addressing. *Cell Syst.* **13** 408–425

Subbiah V 2023 The next generation of evidence-based medicine. *Nat. Med.* **29** 49–58

Tena JJ and Santos-Pereira JM 2021 Topologically associating domains and regulatory landscapes in development, evolution and disease. *Front. Cell Dev. Biol.* **9** 702787

Uversky VN 2019 Intrinsically disordered proteins and their 'mysterious' (meta)physics. *Front. Phys.* **7** 10

Von Hippel PH, Revzin A, Gross CA, *et al*. 1974 Nonspecific DNA binding of genome regulating proteins as a biological control mechanism: 1. The *lac* operon: equilibrium aspects. *Proc. Natl. Acad. Sci. USA* **71** 4808–4812

Waddington CH 1942 Canalization of development and the inheritance of acquired characters. *Nature* **150** 563–565

Waddington CH 1961 *The nature of life* (London: Allen & Unwin)

Wagner A 2015 *Arrival of the fittest: How nature innovates* (New York: Current)

Wagner DE, Weinreb C, Collins ZM, *et al*. 2018 Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360** 981–987

Walsh DM 2015 *Organisms, agency, and evolution* (Cambridge: Cambridge University Press)

Wang RN, Green J, Wang Z, *et al*. 2014 Bone morphogenetic protein (BMP) signaling in development and human diseases. *Genes Dis.* **1** 87–105

Wright PE and Dyson HJ 2015 Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16** 18–29

Zhu R, del Rio-Salgado JM, Garcia-Ojalvo J, *et al.* 2022 Synthetic multistability in mammalian cells. *Science* **375** eabg9765

Corresponding editor: Renee M Borges