

## Opinion piece



CrossMark  
click for updates

**Cite this article:** Ball P. 2016 The problems of biological information. *Phil. Trans. R. Soc. A*

**374:** 20150072.

<http://dx.doi.org/10.1098/rsta.2015.0072>

Accepted: 17 December 2015

One contribution of 21 to a theme issue  
'DNA as information'.

### Subject Areas:

chemical biology, complexity, biophysics,  
statistical physics, computational biology

### Keywords:

DNA, complexity, biological physics, molecular  
biology, information, genomics

### Author for correspondence:

Philip Ball

e-mail: [p.ball@btinternet.com](mailto:p.ball@btinternet.com)

# The problems of biological information

Philip Ball

18 Hillcourt Road, East Dulwich, London SE22 0PE, UK

The discovery of genetic encoding in the DNA molecule, and its mode of translation into protein structures, secured the modern view of biology as an information science. But it remains unclear what kind of information science it is. The all-too-ready analogy with computer programs stored on spools of magnetic tape has been hard to relinquish, even while the complexity of information storage and flow in the cell has become ever more apparent. To understand how life is sustained and evolves through encoding and processing of information, new ideas are now required, within which genetic encoding in DNA seems likely to provide only one part of a much broader and more profound puzzle. In particular, it seems likely that the emerging picture will need to take a more subtle view of causation, context and meaning in the orchestrated, hierarchical processes that make life possible.

## 1. Introduction

One of the most striking features of science in the early twenty-first century is the convergence in the foundations of several core disciplines on the notion of information. In biology, this theme has become prevalent at least since the discovery by Crick and Watson in 1953 of a molecular mechanism for encoding genetic information into the DNA molecule, although it was already implicit in Erwin Schrödinger's *What Is Life?* (1944). Much of the current understanding of genetic mechanisms of cell biology and evolution is now predicated on an image of information transfer among biomolecules, genomes and organisms.

In physics, a unification of ideas from Claude Shannon's information theory with the statistical physics of Maxwell, Boltzmann, Gibbs and others has led to a deeper understanding of concepts such as entropy, irreversibility, the second law of thermodynamics and the 'arrow of time' [1]. Meanwhile, quantum mechanics

is steadily being reframed as a theory of information [2,3]: specifically, in terms of the constraints on what can and cannot be known about quantum systems, and of informatics of the correlations between particle states that distinguish quantum from classical physics.

Even in chemistry, there is a trend towards ideas of programming functions into molecules, imbuing them with a capacity for concerted behaviour that can offer control of matter at scales beyond the molecule [4,5]. On the one hand, this concept derives impetus from the examples supplied by biology. On the other hand, it represents a finer-grained view of the ‘smart materials’ paradigm in which passive substances are being supplanted by fabrics that interact with and respond to their environments in diverse and useful ways, as well as of the notion of ‘programmable matter’ in robotics [6].

An informational perspective has naturally always been a feature of both computer science and neuroscience. But this perspective has become more sophisticated, in line with an appreciation that these two disciplines require more than simplistic parallels: the brain is not a digital computer, artificial intelligence is not the same as ‘making a synthetic brain’, and we do not yet know what form of logic the mind employs.

One could deduce from all of this that an understanding of information processing and transfer in complex systems is the key to a great deal of both fundamental science and technological capability. That might be so. But one might equally see in this apparent convergence nothing more than a continuation of the ages-old tendency to frame scientific ideas in the shape of our current most advanced technology. In the seventeenth and eighteenth centuries, it was common to regard everything from the cosmos to the human body and mind as varieties of clockwork operating according to Newtonian laws of motion. Now that computer and information technologies are the defining mechanisms of our time, are we simply refracting all our core theories through that culturally contingent lens?

That is possible—but it is precisely why considerations of the ‘complexity’ of DNA and genomics can be valuable. Just as it seems to be increasingly clear that simple analogies between the brain and a digital computer are naive and misleading, so too what appeared half a century ago to be an astonishingly good fit between the mechanisms of genetic information storage and transmission and the way these processes are effected in computers is now being revealed as an over-simplification. Genetics surely *is* an information science in some respect, but that does not mean it can be mapped onto systems of Boolean logic, or that the genome can be regarded as an organism’s ‘hard drive’, or that its message is a linear set of symbols that maps transparently onto the phenotypes through which evolution ultimately acts. A complexity viewpoint can offer some inoculation against these idealized pictures, which still, by virtue of their very simplicity, have a deceptive allure in molecular biology.

What will take the place of this simple description? We do not yet know. This theme issue explores some ideas in that direction. Not all of the contributions herein will become bricks in the edifice of a new informational paradigm for life. But some might, and that is exciting.

## 2. The beginnings of information biology

In retrospect, it seems strange that the linear ‘blueprint’ model of genetic information ever took root. In the early days of DNA’s enthronement as the material of inheritance during the 1950s, the stage seemed set for an exploration of just the sort of ideas that this theme issue presents. Norbert Wiener’s cybernetics [7] not only established the notion that organisms can be considered in informational terms as complex systems governed by hierarchical principles of feedback and control, but also restored to a kind of respectability a word essential to understanding biology but long considered taboo in science: teleology. Purpose and design are of course fraught and dangerous words in biology, but there seems to be no avoiding them. Perhaps the truth is that we lack a language that enables us to speak of ‘design’ in an evolutionary context, free from implications of predetermined goals or intelligent formative agencies (let alone from the potential for calculated creationist abuse). Certainly, the gap between the subtleties, complexities and counterintuitive consequences of Darwinian evolution acknowledged by evolutionary biologists,

and the rather simplistic pictures often presented in popular discourses on evolution and genetics, testifies to unease about how best to tell this story, so far as it is currently known [8].

Be that as it may, Wiener's reintroduction of teleology in his 1943 paper 'Behaviour, purpose and teleology' [9] was deliberately provocative [10]. And the Teleological Society that Wiener formed two years later with John von Neumann announced as its mission the understanding of 'how purpose is realised in human and animal conduct'. A 1946 meeting on 'the physics of living matter', attended by von Neumann, seemed to promise that this would be a fruitful arena in the following years. In some ways it was, of course, for Francis Crick seemed to draw on his physicists' training when he elucidated the nature of biological information encoding in the 1950s and 1960s. Among the other early pioneers of 'informational biology' were several other physicists, such as Schrödinger, Max Delbrück, Seymour Benzer—and George Gamow, whose 1955 article 'Information transfer in the living cell' [11] would have fitted well into this theme issue.

But this 'physics of biology' programme stalled, and cybernetics found itself stuck at the qualitative, conceptual level of description—as little more than a source of metaphor. These metaphors were useful, not least for giving Jacques Monod and François Jacob a language for formulating their early theory of gene regulatory networks in the 1960s. But what is required today are models that offer quantitative predictions. That is the test in which the ideas collected herein will ultimately stand or fall.

So this theme issue of *Philosophical Transactions* may be regarded as a re-emergence of a venerable but neglected field. One can speculate about the reasons for why there should be a revitalization now. A possibility is that the low-hanging fruits of the 'DNA revolution' unleashed by Crick and Watson have now been plucked, the last of them being the decoding of the human genome, which was mostly completed by 2001. The technologies that genomics have produced create their own impetus—data-gathering has become an end in itself, and there is now the temptation to imagine that the deep questions will be cracked if only more data are available. But it now looks hard to make much more progress in understanding how genes condition cell behaviour and respond to evolutionary change without a fresh influx of theory.

It was natural in the 1960s to imagine that this understanding could be found in the primary sequence of the genome. It seems unlikely to the point of delusory to imagine that is still the case today. Not only does the genome evidently have higher levels of complexity—most probably governed by the kind of interactive, nonlinear principles Wiener envisaged—but any consideration of 'the complexity of DNA' evidently has to look beyond DNA itself: to the roles of histones, RNA, DNA–protein interactions, organizational processes in the cell that do not require genetic encoding, and interactions between genes, behaviour and environment.

### 3. What is biological information and what role does it play?

If we accept that it makes sense to consider several sciences, from quantum to evolutionary, as problems of information—and even if we also accept that this need not imply an equivalence with binary digital processing in computers—we must then ask to what extent this view is merely metaphorical. It is *as if* there is a 'language of the genes' [12]—one that is duly 'transcribed' and 'translated' in the cell. Or is this more than 'as if'? That, certainly, has been the message of a great deal of thinking in quantum physics and the microscopic basis of thermodynamics. John Wheeler's concept of 'It From Bit' postulates that all physical entities derive their existence from yes/no binary logic applied to measurements and interactions [13]. To put it simply, 'information is physical' [14]—or equivalently, 'the concepts *reality* and *information* cannot be separated from each other' [15]. Indeed, recent experiments have confirmed that information has physical effects: it can be converted to energy [16], and information erasure has an energetic cost [17].

These findings of a fundamental constructive role for information offer a new perspective on the existence of information in biology. There is a school of thought, going back to Hubert Yockey's work on information theory in biology in the 1950s [18], that 'mere' chemistry is not an informational process and cannot therefore offer a self-contained account of the origin of

life. There is a whiff of vitalism here, but that is not to denigrate the idea—rather, it serves to highlight the profound difficulty of understanding how the property of information encoding could have arisen from scratch in the prebiotic world. Of course, we can *design* molecules as information carriers, transmitters and processors, but the very mention of design emphasizes how such things are put in by hand. That idea is linguistically implicit in Barbieri's assertion in this theme issue that living matter is distinct from non-living by virtue of being 'manufactured' matter [19]. This point of view—specifically, the fact that living systems make observable *things* with distinct structures and functions, and do so via a system of coding—leads Barbieri to propose that it is not enough to propose that life is 'chemistry plus information'—we must include 'plus codes' too.

Wills, meanwhile, argues that spontaneous symmetry-breaking transitions are the missing ingredient that allowed a prebiotic system of information coding to emerge [20]. Regardless of whether this can be securely demonstrated (like all origin-of-life ideas, it remains frustratingly hard to test), it has the distinct virtue of encouraging us to cease imagining (if we ever did) that life consists of islands of order surrounded by chaos. The more we know about complex, non-equilibrium systems of all kinds, the more we seem forced to conclude that ordering and organization are a kind of fundamental impulse in matter—not in any mystical sense but as an emergent outcome of the kind of long-ranged correlations that interactions between many components can orchestrate. Such spontaneous self-organization undoubtedly plays a role in life [21,22]; the question is how much can be entrusted to it, and how much the cell needs in the way of specific, bespoke instructions to keep it on course.

Where, then, precisely is the information in a living cell? In traditional views of evolutionary genetics, that question seems to have an obvious answer: it is in the DNA (or at least, in parts of it), since this is what is inherited and as such is the ultimate substrate for evolution. That, however, is surely only part of the answer. (We should also bear in mind that some philosophers and scientists still question whether DNA can be properly said to contain any information at all—for an excellent discussion, see [10].)

For one thing, genomic information is not self-contained within the primary DNA sequence and which then needs simply to be read out like a cassette tape. That has always been clear from the existence of introns and post-translational modification of proteins, but more recently the roles of epigenetic overwriting of genetic messages have also been stressed. Beyond this, it appears likely that some genomic information is encoded via mechanical properties of DNA such as sequence-dependent bending rigidity [23], while more is seemingly regulated by the three-dimensional structure of chromatin, which is able in itself to control gene activity and is actively reorganized for functional reasons [24]. As Wills points out [20], 'DNA sequences are not meaningful except in relation to the operations performed by molecular systems to which they are presented. DNA sequences have no greater 'natural' or 'intrinsic' meaning than sequences of copolymers synthesized randomly in a test-tube, except through their physical and historical embedding in biological systems.'

One way of dealing with such issues is to speak in terms of new levels of encoding: an epigenetic code [25], a DNA repeat-sequence code [26], a chromatin-folding code, even additional codes within the primary sequence itself [27]. It is not yet clear, however, whether the linguistic analogy of coding that is rightly applied to primary gene sequence remains appropriate for these higher levels of informational organization—we do not really yet grasp those aspects of the logic of the cell.

Beyond this, the biochemistry of the cell is not in any case simply a matter of proteins and nucleic acids. That is obvious, but it is often assumed that, because for example the synthesis of fatty acids, hormones, carbohydrates and other classes of molecule is directed by enzymes, their properties and function are somehow wholly genetically encoded. But some of these properties depend on physico-chemical principles that do not need to be transmitted by inheritance, and yet, insofar as they rely on certain types of molecular organization, have a genuine informational aspect. Self-assembly of lipids is spontaneous, and may also be remarkably complex. Even hydration shells often play an active role in biochemical processes, transmitting

information beyond the boundaries of macromolecules or imparting functionally essential dynamical behaviour [28].

It is partly for such reasons that the genome is not a ‘blueprint’ or ‘instruction manual’ for the cell, but rather, a shorthand crib for the many details that cannot otherwise be reliably communicated or assumed between generations. DNA surely encodes some instructions, but not the whole context in which they acquire meaning and causative affect. As Wills [20] puts it here, ‘DNA is being deposed from its role as the Master Molecule to that of an information-carrying servant of quasi-autonomous molecular biological processes’.

## 4. Mechanisms of change

How does evolution on the one hand, and development on the other, make use of biological information? And how are the two related? These are not (yet?) well-posed questions, though Roederer [29] and Koonin [30] here attempt to make them more so. They hinge on the distinction between information and meaning, which is precisely what Shannon and Weaver’s information theory neglects. As Koonin says [30], ‘in order to adequately describe genome function and evolution, the concepts of information theory have to be adapted to incorporate the notion of meaning that is central to biology’. Within Shannon’s theory, information is inversely related to the efficiency of algorithmic compression of a sequence of bits: the most ‘information-rich’ are random. Genetic information is different: in general, coding sequences encode a functional protein sequence, which relates to its folded three-dimensional shape in a manner that is non-trivial, far from transparent, digital to analogue in nature (as remarked on here by Tlustý [31]), and in short, not really understood. It is still not generally possible to infer or predict protein function from the respective nucleotide sequence, for instance, although it is now possible computationally to predict the native fold of some proteins from their amino acid sequence [32], and thence to make inferences from homologies.

What, moreover, is a protein enzyme actually doing in informational, as opposed to strictly biochemical terms? Varn & Crutchfield [33] suggest a way to begin thinking about that question by calling enzymes ‘information catalysts’, akin to (though not identical to) Maxwell’s demon in their ability to use information, drawn in this case from a DNA reservoir, to create a kind of order in the face of thermodynamic obstacles. This, they point out, is not dissimilar to the molecular view of evolution promoted by John Maynard Smith & Eors Szathmáry [34]. However, the ‘meaning’ of a protein can be considered to extend beyond its specific catalytic function: this in itself is of little biological relevance until it is embedded in the respective cascade of biochemical transformations, coordinated in time and space.

Barbieri reminds us that there is nothing particularly new or controversial in an admission of ignorance about the molecular basis of evolution or development [35]. He cites Maynard-Smith, writing in 1986:

It is popular nowadays to say that morphogenesis (that is the development of form) is programmed by the genes. I think that this statement, although in a sense true, is unhelpful. Unless we understand how the program works, the statement gives us a false impression that we understand something when we do not [36, p. 99].

What *has* changed in 30 years is a better recognition of the extent of our ignorance. It was widely thought, when Maynard-Smith was writing, that we did not understand how development is programmed by genes because we had not ‘decoded’ the human genome. Now that we have, it seems increasingly clear that among the many and useful answers to be found in that sequence of bases is not this one, because there is no linear genotype-to-phenotype correspondence. There are more rules to learn, more systems to navigate.

For example, while of course it has been long known that non-coding DNA can have important functional roles, the sometimes heated recent debate [37,38] over the extent of ‘functionality’ in the non-coding genome [39] shows that even rather basic definitions have not yet been agreed



on. There is probably no sharp distinction between ‘meaningful’ (coding or regulatory, say) and ‘meaningless’ (parasitic) sequences. As Koonin argues [30], there is a case for recognizing degrees of ‘fuzziness’ in biological meaning, as well as a state of flux between these degrees. Sequences with fuzzy meaning might, for instance, provide a reservoir of material that can, under certain circumstances, be recruited to provide new functional molecules or regulatory modules. This potentially useful material may therefore be ‘worth’ transcribing to some degree even if it has no obvious immediate functional role. It is for such reasons that issues of meaning become central to the interplay between evolution and development.

The question of why life is ‘evolvable’ at all is more complex than often assumed. It is not obvious *a priori* that small mutational steps should permit adaptation rather than simply inevitable loss of function. Nor is it clear why such a mechanism should permit genuine evolutionary innovation rather than being confined to a sort of timid tinkering with existing functionality. What seems crucial here are the informational features of the genotype–phenotype state-space structure, as elucidated by Andreas Wagner and co-workers [40]. It turns out that, for both DNA and proteins, function can be preserved along whole series of small mutational steps in the primary sequence, which together form a connected network in the space of all possible sequences. Yet at the same time each network is surrounded by many neighbouring ones: a step in the ‘wrong’ direction can take evolution off one network and onto another, with a very different functionality.

A corollary is that biological functionality is highly redundant at several hierarchical levels. Many rather (or even substantially) different protein sequences, for example, can perform the same role. Thus, while the ‘meaning’ of a sequence can to some degree be deduced from homologies, it is not in general the case either that similar protein structures must have similar functions or that similar functions demand similar structures. Indeed, examples are known of structurally very different proteins performing the same function, and structurally similar proteins having different functions [41]. What is more, some proteins have more than one function, and indeed more than one fold (the so-called chameleon sequences) [42].

It is this delicate balance of robust and optimizable functionality with capacity for innovation that makes Darwinian evolution possible. This morphology of the phase space is not unique to proteins and DNA, but applies also to metabolic networks [43] and gene regulatory networks [44]. In fact, it seems to be a general property of complex systems made from permutable elements, and not even unique to biology [45], implying that it is instead an aspect of how complex information is naturally organized—albeit one that is as yet poorly understood.

## 5. Where does causation reside?

A lack of understanding of the ‘logic of life’ [46] leaves many other important questions unanswered in terms of how DNA fulfils its roles in both development and evolution. There is no consensus, for example, on whether the genetic system of life on earth is optimized or unique [47]. Certainly, there is no fundamental obstacle to expanding the genetic code synthetically [48]—but while this can be practically useful in biotechnology, we do not know if it is evolutionarily stable, or whether completely different systems of informational encoding are viable. And while it seems clear today that much of the robustness of living organisms—their ability, for example, to adapt to environmental change without genotypic changes—arises from the properties of gene networks, the rules that govern these features are not known.

A key missing (or at least under-explored) ingredient in this picture is surely dynamics. The structural paradigm has dominated molecular biology ever since the field arose from X-ray crystallography through the work of Bernal, Lonsdale, Astbury, Pauling and culminating in the discoveries of Watson, Crick, Franklin and Wilkins. The broadening of that perspective offered here by Varn & Crutchfield [33] to encompass both orderly and disordered matter is long overdue—not least because the ‘organization’ that exists in biology does not rely on periodicity (as Schrödinger, and earlier Hermann Muller, observed). We now know, for example, that protein functionality does not have to be associated with a specific three-dimensional structure [49]. But

surely one of the key features both of the cell in general and often of the molecular mechanisms that sustain it is that they are *not* static. They are dynamical non-equilibrium structures, and dynamical principles extend all the way from protein–ligand binding (no longer regarded as a geometric lock-and-key) to signalling cascades, metabolism and cell migration and aggregation. Evolution too is evidently a dynamical process, traditionally described in terms of trajectories over fitness landscapes—while acknowledging that both the trajectories and the landscapes are changing in time, and that they are dynamically related.

Walker *et al.* [50] begin to explore that notion here with an analysis of how information flows in a real biological gene network. What distinguishes this network as ‘biological’ seems to be not simply the topology (which is scale-free) but the way in which the collective behaviour is regulated by channelling of information through specific modes, and how this plays a *causative* role.

In my view, this issue of causation will be central to an understanding of biological function, and could elucidate the true role of the genome beyond simplistic blueprint metaphors. Biological information and meaning do not simply arise from the bottom up. In complex systems, higher levels of organization may play an important or even dominant role in causation [51]. In line with this idea, Walker *et al.* [50] propose that the information flow, and causative agency, in living cells can work from the higher organizational levels downward. This runs counter to the bottom-up causation that is familiar in physics (and which is often assumed implicitly in a description of biological behaviour governed by genes). It is, Walker *et al.* suggest, nothing short of a new kind of physics, in which dynamical laws governing the interactions of the components of a system may change with time according to the present global state of the system and its history.

Elsewhere Walker and co-workers have proposed that this same top-down causation may operate in evolution itself [52]—and that the very origin of life was an example of such a transition in causal structure from bottom-up to top-down [53]. The idea that evolution is itself a kind of physical law—an inevitable collective state that emerges in a phase transition from a complexity threshold in the transmission and copying of information—has been suggested previously by Eigen [54].

One can often get the sense that evolutionary and molecular biology today, concerned with the fine details of how evolution gives rise to structure, function and diversity, and with how cells and genes orchestrate their extraordinarily elaborate pathways, have little to gain from the resolution of such abstract questions. But if they are indeed an aspect of the same issues of biological meaning, causation and information that direct the day-to-day functioning of living cells, they are likely to seem harder to ignore. A single meeting on ‘DNA and complexity’ cannot be expected to supply many answers. But it is a good place to start.

**Competing interests.** I declare I have no competing interests.

**Funding.** I received no funding for this study.

## References

1. Parrondo JMR, Horowitz JM, Sagawa T. 2015 Thermodynamics of information. *Nat. Phys.* **11**, 131–139. (doi:10.1038/nphys3230)
2. Brukner C, Zeilinger A. 2003 Information and fundamental elements of the structure of quantum theory. In *Time, quantum and information* (eds L Castell, O Ischebeck), pp. 323–354. Berlin, Germany: Springer.
3. Brassard G. 2005 Is information the key? *Nat. Phys.* **1**, 2–4. (doi:10.1038/nphys134)
4. Lehn JM. 1995 *Supramolecular chemistry*. Weinheim, Germany: VCH.
5. Whitesides GM, Boncheva M. 2002 Beyond molecules: self-assembly of mesoscopic and macroscopic components. *Proc. Natl Acad. Sci. USA* **99**, 4769–4774. (doi:10.1073/pnas.082065899)
6. Goldstein SC, Campbell J, Mowry TC. 2005 Programmable matter. *IEEE Comput.* **38**, 99–101. (doi:10.1109/MC.2005.198)

7. Wiener N. 1948 *Cybernetics: or, control and communication in the animal and the machine*. New York, NY: Technological Press.
8. Ball P. 2013 Celebrate the unknowns. *Nature* **496**, 419–420. (doi:10.1038/496419a)
9. Rosenblueth A, Wiener N, Bigelow J. 1943 Behavior, purpose and teleology. *Philos. Sci.* **10**, 18–24.
10. Cobb M. 2015 *Life's greatest secret: the race to crack the genetic code*. London, UK: Profile.
11. Gamow G. 1955 Information transfer in the living cell. *Sci. Amer.* **193**, 70–78.
12. Jones S. 1993 *The language of the genes*. London, UK: HarperCollins.
13. Wheeler JA. 1991 Sakharov revisited: 'It from Bit'. In *Proc. 1st Int. A. D. Sakharov Memorial Conf. on Physics* (ed. M Man'ko). Commack, NY: Nova Science Publishers.
14. Vedral V. 2010 *Decoding reality: the universe as quantum information*. Oxford, UK: Oxford University Press.
15. Zeilinger A. 2010 *Dance of the photons*. New York, NY: Farrar, Strauss and Giroux.
16. Toyabe S, Sagawa T, Ueda M, Muneyuki E, Sano M. 2010 Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* **6**, 988–992. (doi:10.1038/nphys1821)
17. Bérut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R, Lutz E. 2012 Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* **483**, 187–189. (doi:10.1038/nature10872)
18. Yockey HP. 2000 Origin of life on earth and Shannon's theory of communication. *Comput. Chem.* **24**, 105–123. (doi:10.1016/S0097-8485(99)00050-9)
19. Barbieri M. 2016 What is information? *Phil. Trans. R. Soc. A* **374**, 20150060. (doi:10.1098/rsta.2015.0060)
20. Wills PR. 2016 The generation of meaningful information in molecular systems. *Phil. Trans. R. Soc. A* **374**, 20150066. (doi:10.1098/rsta.2015.0066)
21. Koch AJ, Meinhardt H. 1994 Biological pattern formation: from basic mechanisms to complex structures. *Rev. Mod. Phys.* **66**, 1481. (doi:10.1103/RevModPhys.66.1481)
22. Ball P. 2009 *Shapes*. Oxford, UK: Oxford University Press.
23. Geggier S, Vologodskii A. 2010 Sequence dependence of DNA bending rigidity. *Proc. Natl Acad. Sci. USA* **107**, 15 421–15 426. (doi:10.1073/pnas.1004809107)
24. Horn PJ, Peterson CL. 2002 Chromatin higher order folding—wrapping up transcription. *Science* **297**, 1824–1827. (doi:10.1126/science.1074200)
25. Turner BM. 2007 Defining an epigenetic code. *Nat. Cell Biol.* **9**, 2–6. (doi:10.1038/ncb0107-2)
26. Cournac A, Koszul R, Mozziconacci J. 2016 The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucl. Acids Res.* **44**, 245–255. (doi:10.1093/nar/gkv1292)
27. Fimmel E, Michel CJ, Strüngmann L. 2016 *n*-Nucleotide circular codes in graph theory. *Phil. Trans. R. Soc. A* **374**, 20150058. (doi:10.1098/rsta.2015.0058)
28. Ball P. 2008 Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108. (doi:10.1021/cr068037a)
29. Roederer JG. 2016 Pragmatic information in biology and physics. *Phil. Trans. R. Soc. A* **374**, 20150152. (doi:10.1098/rsta.2015.0152)
30. Koonin EV. 2016 The meaning of biological information. *Phil. Trans. R. Soc. A* **374**, 20150065. (doi:10.1098/rsta.2015.0065)
31. Tlusty T. 2016 The self-referring DNA: a remark on physical and geometrical aspects. *Phil. Trans. R. Soc. A* **374**, 20150070. (doi:10.1098/rsta.2015.0070)
32. Dorn M, Silva MB, Buriol LS, Lamb LC. 2014 Three-dimensional protein structure prediction: methods and computational strategies. *Comput. Biol. Chem.* **53B**, 251–276. (doi:10.1016/j.compbiolchem.2014.10.001)
33. Varn DP, Crutchfield JP. 2016 What did Erwin mean? The physics of information from the materials genomics of aperiodic crystals and water to molecular information catalysts and life. *Phil. Trans. R. Soc. A* **374**, 20150067. (doi:10.1098/rsta.2015.0067)
34. Maynard-Smith J, Szathmáry E. 1998 *The major transitions in evolution*. Oxford, UK: Oxford University Press.
35. Barbieri M. 2016 A new theory of development: the generation of complexity in ontogenesis. *Phil. Trans. R. Soc. A* **374**, 20150148. (doi:10.1098/rsta.2015.0148)
36. Maynard Smith J. 1986 *The problems of biology*. Oxford, UK: Oxford University Press.



37. Eddy SR. 2013 The ENCODE project: missteps overshadowing a success. *Curr. Biol.* **23**, R259–R261. (doi:10.1016/j.cub.2013.03.023)
38. Germain P-L, Ratti E, Boem F. 2014 Junk or functional DNA? ENCODE and the function controversy. *Biol. Philos.* **29**, 807–831. (doi:10.1007/s10539-014-9441-3)
39. ENCODE Project Consortium. 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. (doi:10.1038/nature11247)
40. Wagner A. 2015 *Arrival of the fittest: solving evolution's greatest puzzle*. London, UK: Oneworld.
41. Petsko GA, Ringe D. 2004 *Protein structure and function*, ch. 4. London, UK: New Science Press.
42. Gendoo DMA, Harrison PM. 2011 Discordant and chameleon sequences: their distribution and implications for amyloidogenicity. *Protein Sci.* **20**, 567–579. (doi:10.1002/pro.590)
43. Rodrigues JFM, Wagner A. 2009 Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput. Biol.* **5**, e1000613. (doi:10.1371/journal.pcbi.1000613)
44. Ciliberti S, Martin OC, Wagner A. 2007 Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput. Biol.* **3**, e15. (doi:10.1371/journal.pcbi.0030015)
45. Raman K, Wagner A. 2011 The evolvability of programmable hardware. *J. R. Soc. Interface* **8**, 269–281. (doi:10.1098/rsif.2010.0212)
46. Jacob F. 1973 *The logic of living systems*. (Transl. by BE Spillmann). London, UK: Allen Lane.
47. Jee J, Sundstrom A, Massey SE, Mishra B. 2013 What can information-asymmetric games tell us about the context of Crick's 'frozen accident'? *J. R. Soc. Interface* **10**, 20130614. (doi:10.1098/rsif.2013.0614)
48. Wang L, Xie J, Schultz PG. 2006 Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249. (doi:10.1146/annurev.biophys.35.101105.121507)
49. Oldfield CJ, Dunker AK. 2014 Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584. (doi:10.1146/annurev-biochem-072711-164947)
50. Walker SI, Kim H, Davies PCW. 2016 The informational architecture of the cell. *Phil. Trans. R. Soc. A* **374**, 20150057. (doi:10.1098/rsta.2015.0057)
51. Hoel EP, Albantakis L, Tononi G. 2013 Quantifying causal emergence shows that macro can beat micro. *Proc. Natl Acad. Sci. USA* **110**, 19790–19795. (doi:10.1073/pnas.1314922110)
52. Walker SI, Cisneros L, Davies PCW. 2012 Evolutionary transitions and top-down causation. (<http://www.arxiv.org/abs/1207.4808>)
53. Walker SI, Davies PCW. 2013 The algorithmic origins of life. *J. R. Soc. Interface* **10**, 20120869. (doi:10.1098/rsif.2012.0869)
54. Eigen M. 2013 *From strange simplicity to complex familiarity: a treatise on matter, information, life and thought*. Oxford, UK: Oxford University Press.