Opinion piece

Author for correspondence:
Peter R. Wills
e-mail: p.wills@auckland.ac.nz

DNA as information

Peter R. Wills[1,2,3]

[1]Department of Physics, University of Auckland, PO Box 92019, Auckland 1142, New Zealand
[2]Institut für Biochemie und Molekularbiologie, Universität Hamburg, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany
[3]Department of Biochemistry and Biophysics, University of North Carolina, 120 Mason Farm Road, Chapel Hill, NC 27599-7260, USA

This article reviews contributions to this theme issue covering the topic 'DNA as information' in relation to the structure of DNA, the measure of its information content, the role and meaning of information in biology and the origin of genetic coding as a transition from uninformed to meaningful computational processes in physical systems.

1. Introduction

The biological significance of DNA lies in the role it plays as a carrier of information, especially across generations of reproducing organisms, and within cells as a coded repository of system specification and stability. These roles of DNA do not find any chemical explanation in terms of the average material properties of DNA as an irregular heteropolymer. To understand DNA's biological action, one must go to the detailed molecular level. And then one also fails to find any simple answer in the DNA itself, because single molecules can have vastly different biological effects, covering the entire range of possibilities depending on the molecular biological context, even though they are identical except for the exchange of a particular one of the 10^9 nucleotide moieties of a genome, most such exchanges having very little effect. This drives us immediately to the conclusion that the DNA in organisms functions as *information* [1] and that the internal DNA-dependent dynamics of cells embody functional information processing, that is, *computation* [2]. DNA-based molecular biological computation can be said to control, perhaps even 'direct', the entire panoply of biochemical events occurring in cells.

The obvious way in which information is stored in DNA, as sequences of letters drawn predominantly from the standard four-letter {A, C, G, T} nucleotide alphabet, has been understood since the discovery

of the substance's dual-linear-polymer, base-paired molecular structure and its mode of complementary chain copying [3,4]. However, as soon DNA is represented in these abstract terms, as information comprising a sequence of arbitrary symbols, there is a new theoretical problem. Changes in the DNA sequence of an organism's genome translate in a regular causative way into biological changes in the concrete physical world: such regularities of causation are exploited by genetic engineers when they make calculated informational alterations to an organism's DNA. What is the nature of the causative connection between DNA sequence information, which is an arbitrary abstraction of a material property, and the reality of events in the physical world of molecules embodying the sequence? This information/matter dichotomy is just one of a set of problems that has beset Western philosophy for centuries. To emphasize this point, we refer to the theological ruminations of Aquinas [5] concerning the Christian doctrine of the 'incarnation'. If God became a man, was the man he became (Christ) still truly a man? In similar vein, how can we accept an arbitrarily defined class of abstract sequences as causes of biological events and preserve our analytical view of organisms as causally closed physical systems? While scientists are unlikely to be empowered to perpetrate anything like the terror of the Inquisition in propagating their views as orthodox truth, we should not underestimate the magnitude and intensity of the internecine disputes that underlie discussion of the role of information in biology, such as have led to the separation of the International Society for Biosemiotic Studies and the International Society of Code Biology [6]. Barbieri [7] neatly solves the philosophical problem with his clear and unambiguous identification of *nominable entities*—they embody an irreducible aspect of the natural world, manifest only in biology but as fundamental to reality as basic physical attributes of nature, such as mass, space and time.

The central problem of understanding the role of information in biology arises when attempts are made to go beyond counting the number of the bits of information in a genome and to link genotype with phenotype. DNA information accumulates predominantly through natural selection, a process which is well understood at the molecular level [8]. But knowing that what survives is the fittest does not enlighten one as to the character of the fitness landscape, that is, why, in terms of its internal structure, one organism is fitter than another. Knowledge of what survives gives no insight into how genotypic information is mapped onto the phenotypic characteristics that define the internal factors, as opposed to environmental factors, contributing to an individual's fitness. In other words, how are we to understand the mapping from points in an extremely high-dimensional (DNA) sequence space, entities as abstract as natural numbers, onto the real-world characteristics of organisms whose lives are at stake in the game of evolution played out according to the rules of physics and chemistry? What is fundamental to the dynamic structure of systems in which such a mapping is maintained, in which the *meaning* of the information they store (in DNA) is exquisitely constrained by the system's detailed internal molecular structure? And what features of the material world provide for the spontaneous emergence of structures so remarkably ordered in comparison with what appears to be the bulk of abiotic matter in which no integrated functionality is visible?

The answers to the closely associated questions 'What is information?' and 'What is the origin of life?' provided in this theme issue are quite disparate and although they do not span the gamut of what has been proposed since the discovery of DNA as information, their diversity, similarities and differences are worth exploring. It would be inappropriate for me, as one of the authors involved, to assume the privilege of setting up standards and evaluating colleagues' deliberations. However, without going that far, it is possible, within a context of explicitly stated premises, to describe the relationships between different approaches and to offer some observations concerning the range and relationships of the views which have been expressed. It is the purpose of this short review to compare and contrast the underlying concepts of information and biological processes expressed in the contributions to this theme issue.

2. Measuring the information in DNA

How the information content of DNA is to be defined and measured has been the subject of considerable disputation. Without doubt, the application of Shannon's formula is relevant, but

views concerning the correct way of interpreting and applying the formula differ widely [9–12]. In the tutorial of Adami [13], the ideas of information, uncertainty and entropy are used to relate the material world of physics to the world of human knowledge but we are not told how any of this relates to the information content of DNA. Varn & Crutchfield [14] provide an enlightening treatment of the problem, absolutely rigorous in terms of the foundations of both physics and computation, locating their discussion within the context of the biological necessity for an *aperiodic* crystalline structure in which to store information in nanoscopic matter [1]. The analysis likens DNA to the novel class of *chaotic* crystalline structures, establishing a new equivalence between molecular events in living and non-living systems, especially in respect of the possibility and measure of molecular information processing. Their results are applicable to both biological and artificially constructed systems. Although there is mention of the origin of life, the work circumvents any attempt to define the boundary of the 'living', but the territory left for those who undertake such a task is considerably narrowed. The information theoretic version of the second law of thermodynamics presented by Varn & Crutchfield [14] is of significance to biology yet to be determined by its application, but their seminal insight that '[t]he existence of natural [Maxwell's] Demons with memory (internal states) is a sign that they have been adapted to leverage temporally correlated fluctuations in their environment' nucleates in the exact description of a single system concepts from the theories of molecular evolution, computation and nonlinear irreversible thermodynamics.

Elsewhere, Adami [15] has described how it is that the probability of emergence of replicating biotic systems is increased if monomers are supplied at individual relative rates that are in proportion to their relative abundances in the biotic polymers. This is the same as saying that random typing is more likely to produce meaningful sequences of letters if the probability of typing letters matches the frequency of occurrence of the letters in the language of choice. However, Adami [15] links the biological notion of selective adaptation to his description of an integrated system of functional polymers with an overall monomer composition matching that of the environment. Restricting consideration to the case of natural selection among molecular self-replicators constrained by the supply of monomers should provide opportunity for a clear demonstration of the results presented by Varn and Crutchfield, perhaps making an explicit connection between thermodynamics and the theory of evolution, such as was left incomplete by Eigen [8] and recently considered by England [16].

Koonin [17] takes the view that biological information is 'effectively orthogonal' to Shannon information, because it has to do with meaning, not the statistical distribution of symbols in a sequence—the biological meaning of sequences can be found only through the alignment of homologous sequences, not by examining individual sequences. No specification of how homology can be rigorously determined is provided (it can only be assumed that bioinformatic techniques for aligning sequences are considered to be adequate) but meaning can be measured through the 'vertical' comparison of aligned sequences from different sources, rather than the 'horizontal' comparison of sites along a single sequence. This meaning is seen to be transferred by DNA exchange between genomes during evolution. Koonin identifies a quantity he calls the 'information density' as a measure of the meaning of a DNA sequence. It corresponds to the average, across an alignment, of the single-sequence-position deviation from randomness. However, in the end, Koonin completely relativizes his information theoretic measure of biological meaning by pointing out that the question 'meaning for whom?' is answered through the range of orthologous sequences chosen for the alignment in relation to which information density is calculated.

3. Systemic information processing

The essence of computation is information processing, and the essence of biological information processing is control of the molecular events inside a cell. Thus, Walker *et al.* [18], taking fission of a complete single-celled yeast cell as an example, locate the special character of biological systems—the connection between information and causation—in the 'informational architecture'

of the cell, the spatio-temporal structure of the transformation of information during biological processes. The information theoretic analysis they present is rigorous and the conclusion reached is that 'biology is distinguished from physics . . . in how the flow of information directs the execution of function'. They describe this in terms of the informational architecture interacting with the system's causal structure, which is construed to be physico-chemical. While the analysis is intended to apply to any level of biological organization, including DNA, there is no exposition of how the connection works at the level of nucleotide sequences and chemical reactions. However, the implicit genetic control of the individual processes of the yeast fission system makes the study relevant to an understanding of exactly how it is that quantities of functional information, originating in a DNA repository and corresponding to dynamically constrained distributions of alternative states, can operate to govern the whole-system behaviour.

Roederer [19] also characterizes biological systems in terms of a causal connection between informational patterns and events in the material world. Biological information is *pragmatic*, its importance lying not in the quantity of it but in what its effect in any system. For example, in the absence of deep epigenetic factors, there is an essentially univocal correspondence between the linear pattern of DNA bases presented to cells of a certain species and the ensuing complex of molecular biological dynamic changes that take place. However, pragmatic information cannot be measured, because it represents a correspondence between a pattern and a change; and thus it is highly context-dependent. This idea of pragmatic information is very close to what others like Koonin [17] and Wills [20] would refer to as the *meaning* of information (e.g. the pattern of nucleotide bases in a genome). As Roederer [19] states 'a pattern all by itself has no meaning', which is echoed in the ansatz of Wills [21]: 'Any body of information can be given any meaning whatsoever, by creating a device that functions as an interpreter to deliver the specified meaning upon reception of that information'. Thus, both Roederer [19] and Wills [21] imply the possibility of epigenetically defined phenotypes, such as the distinguishable strains of yeast that breed true in the same environment even though they have identical genomes [22], a problem that was set aside in the first characterization of heritable information in nanoscopic structures [1].

Like Varn & Crutchfield [14], Wills [20] focuses on the linear, aperiodic crystalline structure of DNA as the medium for the static physical instantiation of biological information, a body of which can constitute the complete, heritable 'specification for the construction and maintenance of an entire organism' [1]. His enquiry delves into how molecular systems can self-organize to fuse a union between the DNA sequence information they contain and the internal molecular componentry that cooperatively and self-referentially generates meaning from the information. The investigation considers the machinery of translation, the system for executing the rules of the genetic code, as an example of a molecular biological interpreter, representative of a system of functional computation that is fundamental to all biological systems. It is argued that the principle of natural selection does not alone account for the evolutionary accumulation of information in DNA [8]. Rather, equal emphasis should be given to processes of *epigenesis*, whereby selective advantage is conferred on genetic sequences by virtue of their coincidental occurrence with new interpretations of them, interpretations that simultaneously emerge, together with the selected information, as a result of functional self-organization within the system. Neither information nor function is given causative precedence. This disruption of the neo-Darwinian, selectionist view of evolution impacts on the adequacy of the Central Dogma's explanation of DNA as information.

4. Central Dogma; origins and regularities of coding

Tlusty [23] looks beyond the Central Dogma's rigid view of information flowing one way from DNA to protein [24] and shows how it is possible to 'close the loop' in the transformations of information taking place in a functioning cellular system. Commoner [25] put himself completely outside the mainstream of molecular biological thinking when he proposed such a project, but a half century later Tlusty [23] is a pioneer in efforts to give a formal description of the

natural foundations, we might even say 'language', of systems biology. Just as the invention of the term 'nominable entities' [7] has encapsulated in two words what many have long thought concerning the status of information in biological systems, or the definitive analysis of England [16] has settled many vaguely framed questions concerning the chemical orderliness of functional biological systems and their internal production of entropy, so too does the work of Tlusty [23] show us something very basic about how information inherited in DNA flows through a pre-existing maze of intricately connected, interacting processes to play its role in keeping cells alive and functioning. A genome-wide application of the methodology of Tlusty [23] would, in principle, give an elementary picture of a cell's 'informational architecture' as Walker *et al.* [18] define it, but then new tools would be needed to make sense of the blueprint that emerged. And one suspects that the picture would always remain incomplete, as every new detail of the epigenetic and other system factors affecting the interpretation of DNA information, like those described by Bartholomäus *et al.* [26], Paci *et al.* [27] or Seligmann [28], came to light.

Both Barbieri [7] and Wills [20] seek a deeper understanding of the character of biological information in the historical origin of genetic coding. Both authors take the emergence of coding as an essential element of the transition from abiotic to biotic chemistry. For Barbieri, the transition across the boundary entails the appearance of *nominable entities*, which can be described only by naming the order of their components, that is, by specifying a pattern of information. This property qualifies them to be designated as manufactured *artefacts*, in stark contrast to all the other molecules in the universe, which are 'spontaneous'. The first genes and proteins were spontaneous, but some molecules somehow took on the status of *machines*, functioning as 'bondmakers' and 'copymakers', in turn conferring on some genes the status of being pattern-preserving, information-bearing *templates*. This enabled the appearance of the first artefacts, molecules whose components were ordered by pre-existing information. However, before the emergence of the genetic code, none of the template-maintained information had *meaning* and it is on this basis that Barbieri [7] proposes his *code paradigm* of life as 'chemistry + information + codes'. Elsewhere, Barbieri [29] attributes the emergence of accurate, unambiguous coding to the evolution of ribosomal proteins, but comprehensive bioinformatic analyses [30–32] point to the precedence of specifically functional aminoacyl-tRNA synthetase (aaRS) proteins in the processes of code expansion.

Although it defines DNA information in a way that is barely distinguishable from many others, especially that of Barbieri [7], the analysis of the origin of coding presented by Wills [20] has a completely different theoretical foundation. It starts with considerations of the dynamics of autocatalytic networks of polymers and the chicken–egg problem of the aaRS coding enzymes being needed for their own synthesis. Instabilities connecting alternative solutions to the dynamic equations describing the elementary chemical processes of gene replication and translation provide the decisive insight into how coding can arise from random peptide synthesis as a result of relatively simple systems undergoing self-organizing thermodynamic transitions. The need to preserve, through the process of natural selection, the genetic information coding the aaRSs demands that the products of genetic replication and translation be colocalized. In the absence of compartmentalized proto-cells that conveniently reproduce their entire contents in the right proportions as living cells do, Turing reaction–diffusion coupling provides an elementary mechanism for polymeric genotypes and phenotypes to maintain colocalization and stave off the caustic effects of computational errors. As a result of dynamic transitions described as 'quasi-species bifurcations' in such systems, coding ambiguity is reduced in parallel with the progressive accumulation of genetic information sufficient for progressively more complex populations of aaRS 'statistical proteins' to specify themselves with expanded codes of increasing precision (reduced error). Thus, Wills [20] claims that the problem of expanding code specificity raised by Barbieri [29] was effectively solved a decade and a half ago [33], that solution having since been carefully elaborated, most recently spawning a novel bioinformatic analysis of the deep co-phylogenies of functional aaRS structures [34]. He suggests that the regularities and apparent optimality of the genetic code are inevitable consequences of the originary mechanism he proposes.

Gonzalez *et al.* [35] consider how arithmetical representations of the map from trinucleotide codons to amino acids can potentially reveal hidden correlations and symmetries that are not evident in normal tabular representations of the genetic code. Each codon *C* is represented by a sequence of six binary digits $d_k \in \{1, 0 : k = 1, 6\}$ and then transformed into an integer N_C through a set of weights w_k by arithmetic summation: $N_C = \sum w_k d_k$. If the weights increase more slowly than the conventional power relation $w_k = 2^{k-1}$, then groups of one or more degenerate codons map onto the same integer. Through judicious choice of the weights, the numbers of degenerate codons mapping onto the integers in the range of the mapping $\mathbf{C} \to \mathbf{N}_C$ can be made to correspond to the genetic coding degeneracies found in organisms or organelles. It is salient that the weights defined by the mapping indicate the relative importance of the corresponding bits in determining which amino acid is specified by a codon, but the individual bits generally bear a poor relation to simple distinguishing properties (e.g. purine or pyrimidine) of the base occurring at a particular codon position. This means that correspondences between properties of bases at more than one position must be invoked to give a full explanation of the pattern of degeneracies seen in any genetic code. Thus, the mapping provides a metric for the detection and analysis of intra-codon short-range correlations in coding sequences. It turns out that the mitochondrial code can be represented by a more symmetric pattern of weights than the standard code, indicating that it is fundamentally less differentiated, probably displaying a pattern closer to a more primitive code from an earlier epoch of molecular biological information processing. The closely related work of Fimmel *et al.* [36] extends the analysis to the influence of the first and third bases of codons over their neighbouring bases in consecutive codons. This is consistent with the suggestion of Di Giulio [37] that the first mRNAs coevolved with peptidated RNAs that catalysed their own synthesis, using base-pairing patterns that were more than three bases long. Thus, it seems that dinucleotides play a direct role in the translation of genetic information as well as in the DNA methylation processes involved in epigenetic signalling [27]. Results from this entire body of work should assist in discovering details of the molecular mechanisms, whereby codons are functionally distinguished, especially through aaRS enzymes' tRNA-selectivity and the ribosomal frame-keeping ratchet. A weighted bit representation of tRNA sequence elements has already been used in an empirical approach through which major molecular-level chemical factors involved in the aaRSs' matching of codons to amino acids have been successfully elucidated [38].

5. Summative conclusion

The topic 'DNA as information' focuses attention on biological information as it can be stored statically in molecular structures. But the essence of information in biology is its dynamic transfer into different forms and the effects of such transfers. Nowhere is that more evident than in the contribution of Walker *et al.* [18], who analyse the consequences of information flows and processing in molecular biological control systems. Their methodology is more reminiscent of the thermodynamic approach of England [16] than others whose work has a direct connection to the Central Dogma [24]. However, we have come a long way from the simple maxim 'DNA makes RNA makes protein', to the extent that it seems unremarkable that Tlusty [23] should ascribe the functional effect of a transcriptional factor binding to DNA as a sequence-information mapping. And this is just one example of countless functional interactions taking place in every cell. Every new discovery of a biological macromolecule or functional effect [26,27] modifies the estimate of the information content of the DNA of the species concerned and other members of its clade. Koonin [17] has proposed a way of quantifying the information in DNA relative to the breadth of the selected clade, giving us a view of how the meaning of information in DNA continually evolves as a result of mutation and gene transfer. Roederer [19] considers the general consequences of some pre-existing molecular pattern influencing the path of physical processes, emphasizing the need for what amounts to 'recognition' of information for it to have meaning in biological contexts. The origin of such processes and their consequences is the focus of the contributions from Barbieri [7] and Wills [20], both of whom take the recognition and matching of codons and amino acids, genetic coding, to be a defining feature of biology and the origin of life.

However, as the work of Varn & Crutchfield [14] shows, there is a much broader class of structures that could potentially mimic the function of 'DNA as information' in molecular biological-like systems. And the work of Gonzalez *et al.* [35] and Fimmel *et al.* [36] demonstrates that the one example we have of meaningful information stored in molecular structures, that is, life on this planet, relies on a system of interpretation that is intricately bound to exquisite details of both the historical process through which it has emerged and the perhaps diverse possibilities, consistent with the laws of physics and chemistry, of self-referential molecular structures and functionalities coevolving. It might be possible, in completely different environments, for systems of nanoscopic processes to bootstrap themselves into existence and evolve as a result of their association with a colocalized repository of non-DNA information, which they manage to interpret as a programme for their construction through a network of processes not involving a simple translation step. The possibility of creating artificial systems of that sort is certainly being explored [39].

References

1. Schrödinger E. 1944 *What is life?* Cambridge, UK: Cambridge University Press.
2. Von Neumann J. 1949 Theory and organization of complicated automata. In *Theory of self-reproducing automata* (ed. AW Burks), pp. 29–87. Urbana, IL: University of Illinois Press.
3. Watson JD, Crick FHC. 1953 Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738. (doi:10.1038/171737a0)
4. Watson JD, Crick FHC. 1953 Genetical implications of the structure of deoxyribose nucleic acid. *Nature* **171**, 964–967. (doi:10.1038/171964b0)
5. Aquinas T. ca 1265 Summa theologica. Third Part, QQ[2]. See http://www.sacred-texts.com/chr/aquinas/summa/sum453.htm.
6. Barbieri M. 2014 From biosemiotics to code biology. *Biol. Theory* **9**, 239–249. (doi:10.1007/s13752-013-0155-6)
7. 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)
8. Eigen M. 1971 Self-organisation of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–532. (doi:10.1007/BF00623322)
9. Battail G. 2014 *Information and life*. Berlin, Germany: Springer.
10. Muller S. 2007 *Asymmetry*. Berlin, Germany: Springer.
11. Schneider TD. 2010 A brief review of molecular information theory. *Nano Commun. Netw.* **1**, 173–180. (doi:10.1016/j.nancom.2010.09.002)
12. Yockey HP. 1992 *Information theory and molecular biology*. Cambridge, UK: Cambridge University Press.
13. Adami C. 2016 What is information? *Phil. Trans. R. Soc. A* **374**, 20150230. (doi:10.1098/rsta.2015.0230)
14. 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)
15. Adami C. 2015 Information-theoretic considerations concerning the origin of life. *Orig. Life Evol. Biosph.* **45**, 309–317. (doi:10.1007/s11084-015-9439-0)
16. England JL. 2014 Statistical physics of self-replication. *J. Chem. Phys.* **139**, 121923. (doi:10.1063/1.4818538)
17. Koonin EV. 2016 The meaning of biological information. *Phil. Trans. R. Soc. A* **374**, 20150065. (doi:10.1098/rsta.2015.0065)
18. 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)
19. Roederer JG. 2016 Pragmatic information in biology and physics. *Phil. Trans. R. Soc. A* **374**, 20150152. (doi:10.1098/rsta.2015.0152)

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. Wills PR. 2014 Genetic information, physical interpreters and thermodynamics; the material-informatic basis of biosemiosis. *Biosemiotics* **7**, 141–165. (doi:10.1007/s12304-013-9196-2)

22. Wickner RB, Shewmaker FP, Bateman DA, Edskes HK, Gorkovskiy A, Dayani Y, Bezsonov EE. 2015 Yeast prions: structure, biology, and prion-handling systems. *Microbiol. Mol. Biol. Rev.* **79**, 1–17. (doi:10.1128/MMBR.00041-14)

23. Tlusty T. 2016 Self-referring DNA and protein: a remark on physical and geometrical apects. *Phil. Trans. R. Soc. A* **374**, 20150070. (doi:10.1098/rsta.2015.0070)

24. Crick FHC. 1958 On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163.

25. Commoner B. 1964 Deoxyribonucleic acid and the molecular basis of self-duplication. *Nature* **203**, 486–491. (doi:10.1038/203486a0)

26. Bartholomäus A, Fedyunin I, Feist P, Sin C, Zhang G, Valleriani A, Ignatova Z. 2016 Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Phil. Trans. R. Soc. A* **374**, 20150069. (doi:10.1098/rsta.2015.0069)

27. Paci G, Cristadoro G, Monti B, Lenci M, Degli Esposti M, Castellani GC, Remondini D. 2016 Characterization of DNA methylation as a function of biological complexity via dinucleotide inter-distances. *Phil. Trans. R. Soc. A* **374**, 20150227. (doi:10.1098/rsta.2015.0227)

28. Seligmann H. 2015 Codon expansion and systematic transcriptional deletions produce tetra-, pentacoded mitochondrial peptides. *J. Theor. Biol.* **386**, 154–165. (doi:10.1016/j.jtbi.2015.09.030)

29. Barbieri M. 2015 Evolution of the genetic code: the ribosome-oriented model. *Biol. Theory* **10**, 301–310. (doi:10.1007/s13752-015-0225-z)

30. Caetano-Anollés G, Wang M, Caetano-Anollés D. 2013 Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. *PLoS ONE* **8**, e72225. (doi:10.1371/journal.pone.0072225)

31. Harish A, Caetano-Anollés G. 2012 Ribosomal history reveals origins of modern protein synthesis. *PLoS ONE* **7**, e32776. (doi:10.1371/journal.pone.0032776)

32. Petrov AS *et al.* 2015 History of the ribosome and the origin of translation. *Proc. Natl Acad. Sci. USA* **112**, 15 396–15 401. (doi:10.1073/pnas.1509761112)

33. Wills PR. 2001 Autocatalysis, information and coding. *BioSystems* **60**, 49–57. (doi:10.1016/S0303-2647(01)00107-1)

34. Wills PR, Nieselt K, McCaskill JS. 2015 Emergence of coding and its specificity as a physico-informatic problem. *Orig. Life Evol. Biosph.* **45**, 249–255. (doi:10.1007/s11084-015-9434-5)

35. Gonzalez DL, Giannerini S, Rosa R. 2016 The non-power model of the genetic code: a paradigm for interpreting genomic information. *Phil. Trans. R. Soc. A* **374**, 20150062. (doi:10.1098/rsta.2015.0062)

36. 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)

37. Di Giulio M. 2015 A model for the origin of the first mRNAs. *J. Mol. Evol.* **81**, 10–17. (doi:10.1007/s00239-015-9691-y)

38. Carter Jr CW, Wolfenden R. 2015 tRNA acceptor stem and anticodon bases form independent codes related to protein folding. *Proc. Natl Acad. Sci. USA* **112**, 7489–7494. (doi:10.1073/pnas.1507569112)

39. McCaskill JS *et al.* 2012 Microscale chemically reactive electronic agents. *Int. J. Unconv. Comp.* **8**, 289–299.