Goethe's dream

Challenges and opportunities for synthetic biology

Andrés Moya, Natalio Krasnogor, Juli Peretó & Amparo Latorre

The German philosopher Immanuel Kant (1724–1804) defined living organisms as objects with an intrinsic purpose, which are self-organized in such a way that every part is a function of the whole and the whole is a function of every part, and in which “nothing is for nothing”. Kant already anticipated the tension between agency and structure, and between forward and backward causation. He also perceived living beings as entities that, being extremely complex, are not amenable to descriptions based on laws that are similar to the fundamental laws of physics: “There will never be a Newton of a grass blade,” he wrote. Less metaphorically, Kant believed that science would not be able to understand living entities by focusing exclusively on their component parts, and, therefore, that it would never be able to explain a whole organism completely and exhaustively.

The history of biological research can be regarded as an attempt to prove Kant wrong

The history of biological research can be regarded as an attempt to prove Kant wrong. We can mainly distinguish two major strategies to achieve this, analytical and synthetic, which in turn have determined how biologists have tried to understand what is a living being. Analytical biology, or the so-called reductionist approach to biology, has mainly focused on the study of the individual components of living organisms. Its last incarnation, molecular biology, has been extraordinarily successful at producing a deluge of data on the molecular mechanisms underpinning life processes and has resulted in tremendous advances in the explanation of life workings (Morange, 2005).

There is a complementary—sometimes antagonistic—tradition in biology, the synthetic tradition, which states that the essence of a living entity cannot be understood by merely studying its parts (Keller, 2002; Peretó & Català, 2007). The German poet and naturalist Johann Wolfgang von Goethe (1749–1832) gave voice to the holistic perspective when he asserted that “what a living being is, its essence can be split up in its elements, but it won't be possible to go back and recompose the object to bring life back to it.” Goethe recognized the contribution that the analytical view has made to our knowledge, but asserted that it would not be sufficient to understand what is life. By contrast, the standing problem with the synthetic view is that, despite its ontologically correct emphasis on attempting to understand a living being without deconstructing it, it lacks a conceptual framework and associated methodological tools to study living beings as complete entities. In an attempt to give credit to the synthetic approach, the German-born American evolutionary biologist Ernst Mayr (1904–2005) stated that notions such as *entelechy* or *élan vital* were introduced to overcome the Cartesian interpretation of living beings as machines, and to stress that the interaction of components is as important as the components themselves in order to understand a living entity (Mayr, 2002).

It is also worth noting that the conceptual and methodological advances made under the analytical view unveiled such a detailed knowledge of biological parts that they created a new synthetic view of biology, which we call ‘synthetic view one’. By way of example, it attempts to synthesize either a protocell or some of its

Computing is an appropriate formal language to describe and reproduce biological phenomena

putative components, or to create a minimal cell by ‘slimming down’ natural cells to minimal functions. Protocell research (Mansy *et al*, 2008; Szostak *et al*, 2001; Rasmussen *et al*, 2008; Pereira de Souza *et al*, 2009) includes fully synthetic systems from abiotic chemical components and engineered vesicles that incorporate some minimal biological components such as proteins, ribosomes, certain enzymes, nucleic acids and other cell extracts (Luisi *et al*, 2006). By contrast, the ‘top-down’ strategy uses a minimal cell as a chassis on which to implement new functions by adding genes for metabolic and regulatory networks (Gil *et al*, 2004; Moya *et al*, 2009; Andrianantoandro *et al*, 2006). This approach is at the core of the engineering view of synthetic biology, which aims to create biological devices that are standardized and show a predictable behaviour, regardless of the biological chassis on which they are implemented (Endy, 2005). Eventually, it might be possible to synthesize a complete genome (Gibson *et al*, 2008) and transplant it into a cellular chassis (Lartigue *et al*, 2007).

However, another concept, ‘synthetic view two’, is probably better suited to capture the complexity of biological phenomena. This view considers biological phenomena as systems of interacting components that can be analysed and simulated with increasing detail at both the level of components and the level of the whole system. Moreover, this view does not exclude emergent properties within the system (Serrano, 2007).

Computing is an appropriate formal language to describe and reproduce biological phenomena. We realize that this statement might be perceived as controversial because quantitative modelling is difficult, especially when the bar is perhaps set too high in contraposition with the success that the physical sciences have had in describing the physical world. More specifically, unlike the remarkable success of calculus as a formal language by which to express, model and derive knowledge about the physical world, biology has had, until now, no appropriate language to characterize formally the complex phenomena of living organisms.

The macroscopic continuous and deterministic approach based on ordinary differential equations is the most widely used methodology in cellular modelling within systems and synthetic biology. Although ordinary differential equations have been successfully applied in different systems, two key assumptions of this approach,

namely continuity and determinism, are not always fulfilled. In many systems, the number of particles of the reacting species is low and the reactions involved are slow. In such cases, the previous assumptions are invalid and mesoscopic; discrete and stochastic approaches, which are closely related to the tools frequently used within computer science, would seem to be more suitable (Szallasi *et al*, 2006; di Ventura *et al*, 2006). The latter approaches have been implemented in different computational frameworks ranging from rewriting systems such as P systems (Romero-Campero & Pérez-Jiménez, 2008a, 2008b) to Petri nets (Goss & Peccoud, 1998) and process algebra (Regev & Shapiro, 2002). In these computational formalisms, biological entities such as signalling molecules, transcription factors, cell membranes and their interactions are explicitly modelled more akin to reactive systems (Sadot *et al*, 2008), rather than by approximating the rate of change for molecular

concentrations as in 'calculus-based' modelling. Indeed, these modelling approaches are being recognized as perhaps an 'executable biology' metalanguage with which detailed and rigorous analysis of biological systems could be done (Fisher & Henzinger, 2007).

Regardless of whether, in the long run, the scientific community settles for a continuous or an executable 'language' when attempting to model synthetic-biology systems, the fact remains that there are limits to what could be formally—that is, mathematically or computationally—proven or demonstrated. These limitations will, sooner or later, reintroduce into synthetic biology the complexity that it is trying to avoid in the form of unaccounted emergent phenomena.

Theoretical computer science, logic and mathematics have profoundly changed the way in which we see the world and our place within it as intelligent beings. Among other works, Kurt Gödel's

incompleteness theorem (Gödel, 1931), Alan Turing's halting problem (Turing, 1936), and Gregory Chaitin's omega number (Chaitin, 1974) all demonstrate that there are limits to what can be formally known and proven to be true. Indeed, for any given formal system, there are statements that cannot be shown to be true or false; in other cases, statements are true or false for no specific reason—that is, we cannot derive a proof. This undecidability barrier has been often used to 'show' that a computer cannot be intelligent, in a human sense, or to argue that brains (Penrose, 1989) and bacteria (Ben-Jacob, 1998; Danchin, 2009) must apply some yet unknown physics to perform computations that would apparently operate beyond the Gödel–Turing–Chaiting (GTC) limit—which represents a limit to what we can know regardless of the amount of time and space that we are willing to invest in order to achieve knowledge.

The other practical limitation to what we can know is intractability (Cook, 1971), which is perhaps more ubiquitous, as it states that certain problems that could be solved in principle actually take far too long to solve—unless one relaxes some conditions—and therefore remain either unsolved or only partly solved. Although some philosophical arguments have been made about whether the GTC limit and intractability pose an insurmountable barrier for systems and synthetic biology, we can gain more interesting and useful insights by considering their practical impact.

...biology has had, until now, no appropriate language to characterize formally the complex phenomena of living organisms

For the sake of argument, let us summarize the goal of synthetic biology as the creation of synthetic organisms by the application of modelling and/or engineering techniques of any kind. Let us also assume that we are interested in so-called white-box models—that is, we know the contents and operations. A legitimate question would be "How do we know if this is a correct model?" or, to put it another way "Can we rely on the model to synthesize the desired organism?" The answer depends on what we mean by 'correct'. A correct model could be one that is computed—that is, simulated—fast enough to be of practical interest or one that is slower

but more accurate. However, there are—as mentioned above—intrinsic limitations to how much faster one can get and the price that is paid in accuracy. By contrast, it is likely that experimental wet-laboratory errors will be larger than those accrued during a simulation; hence, even if we, in principle, could run a simulation on a yet-to-be-invented supercomputer, for what would that be useful? We could ask whether a model is reliable—that is, whether it has been formally verified to be accurate. Here again, although great progress has been made in the formal verification of software, there are fundamental limitations of the two types mentioned previously: undecidability and intractability.

It is safe to say that, for the foreseeable future, formal verification of computer models will be limited to relatively small systems. Having said this, intractability has not, historically, been an impenetrable barrier. We are now able routinely to solve specific instances of intractable problems, such as the travelling salesman and graph-colouring problems, even for millions of nodes to within small margins from the optimal solution. There is a more insidious meaning, however, to the question of "How do we know if this is a correct model?", which a biologist would more readily associate with—namely, the suitability of the model to capture all extant knowledge about a biological system. That is to say, how do we compare the behaviour of a computer model against that of a biological system? To address this question, we must, again, make two ontological assumptions.

Let us initially assume that biological systems and computer models are not intrinsically different. One might hastily conclude that, as both the biological entity and the model are not fundamentally different in kind, they can be directly compared. However, even for systems that are defined within the same formalism, such as a computer code, there are still fundamental limitations to what we can say about their behavioural equality.

Now let us assume that biological and computer systems are intrinsically different. In this case, one might conclude that there is no efficient method to assess the adequacy of a computer model to simulate a biological entity. Again, it would seem to us that this conclusion would be premature.

Ignoring either assumption, there is a pragmatic methodology that could, in principle, tackle the issue of model adequacy.

Harel (2005) proposes a modified Turing-like test for incrementally refining-while-testing computer models of biological systems. According to the Harel protocol, all available knowledge of a biological system, for instance, *Caenorhabditis elegans*, is included in the model, which will be deemed successful if, to the eye of an expert, its behaviour is indistinguishable from that of a true *C. elegans*. If a given experiment reports a discrepancy between the behaviour of the nematode and the model, then that discrepancy should be captured in the model, thereby over time resulting in better models—and ultimately in better biological understanding. The proposed methodology relies on no assumption about whether living organisms can be accurately modelled on a computer; hence, in principle, it is a valid protocol to follow for improving systems-biology models regardless of any putative fundamental limitation given by the GTC limit or intractability limit.

...how do we compare the behaviour of a computer model against that of a biological system?

In a similar vein, and pursuing the same goal as Harel but within the context of synthetic rather than systems biology, Cronin *et al* (2006) have proposed the use of real cells within a Turing-like test to measure whether artificially engineered protocells behave similarly to their authentic counterparts. Interestingly, this would be both a practical and a valid test for synthetic life, even if the second ontological assumption—that computer and biological systems are fundamentally different—turned out to be true; here, 'nature' itself would be doing the 'simulation' instead of a computer program as in the Harel proposal.

In any case, we think that GTC limits and intractability lead to 'practical' emergent properties when modelling and implementing synthetic systems. Indeed, the concept of 'emergent properties' is fundamental to biology: most biological characteristics and evolutionary advances are not predictable, owing either to intractability or to formal undecidability. In summary, emergent phenomena will, sooner rather than later, appear within any sufficiently complex system, which calls for rethinking the way in which methodologies in synthetic biology are conceptualized and

developed. We might then define a new biological system by adding new rules that integrate the emergent features.

Synthetic biology has been interpreted as an engineering discipline (Endy, 2005; Andrianantoandro *et al*, 2006), whereby both the whole cell and any natural or artificial cellular components should be standardized—that is, they should show a predictable and controllable behaviour. The emphasis on control is of great relevance because it means that any man-made biological device should behave always as expected in any suitable environment. To avoid uncertainties, we need to apply two levels of quality control: first, by exploring all imaginable environmental circumstances during the design stage; and second, when the device is put into a biological chassis and released. It is probably much easier to control simple biological devices than living cells, and, within cells, much easier to control a single minimal cell than a complex one. Some proponents of the engineering view of synthetic biology therefore comment that we should not put too much emphasis on building artificial cells, but instead should focus on engineering simple devices with many possible biological applications (Endy, 2005; Andrianantoandro *et al*, 2006).

...the concept of 'emergent properties' is fundamental to biology: most biological characteristics and evolutionary advances are not predictable...

Synthetic biology can also be regarded as an applied methodology for the creation of biological systems from which we are gaining knowledge. Complementary to the engineering view, this approach embraces complex phenomena and emergent properties. Advances in all areas of molecular and computational biology, coupled with recent developments in network and graphs theory, allow us to simulate cellular behaviour, manipulate the cell and observe the outcomes of such interventions. This view, which we call the systems-biology approach to synthetic biology (Serrano, 2007), is probably better suited to meet the expectations of professional biologists. It is also under the umbrella of the second synthetic view mentioned earlier and could realize Goethe's idea of understanding life as a whole. However, this approach not

only deals with basic science or the first principles of living beings, but is also a highly applied programme.

The history of biology could be interpreted as a permanent dialectic battle between two confronting views on how to deal with and comprehend living entities: integrative versus analytical. The first view promotes a cohesive approach, whereby parts are integrated into wholes, as the only way to comprehend the true nature of a living entity. The second view approaches life by concentrating mainly on its components. This analytical approach has led to the development of powerful conceptual and methodological tools to dissect the components of living beings and their complex interactions. By contrast, with the advent of the network theory and advances in computation, a more comprehensive and context-sensitive view of biology is emerging. This synthetic biology, which is a combination of both approaches, might represent a fundamental step towards the realization of Goethe's idea.

We are confident about the feasibility of building a living being, or a part of it, whether it is one that already exists or a non-natural entity. This particular view of synthetic biology is likely to enable important applications in areas as different as medicine, energy production and the environment. It requires a serious reflection on the corresponding responsibilities and on the consequences of these new entities that we humans might create. In addition, synthetic biology demands a philosophical assessment of the panoply of futures that could be realized by humankind.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community Seventh Framework Programme (FP7/2007-2013) under grant agreement number 212894, TARPOL consortia. N.K. and A.M. thank the Biotechnology and Biological Sciences Research Council for funding project SynBioNT: A Synthetic Biology Network for Modelling and Programming Cell-Cell Interactions (BB/F01855X/1).

REFERENCES

- Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* **2**: 2006.0028
- Ben-Jacob E (1998) Bacterial wisdom, Godel's theorem and creative genomic webs. *Physica A* **48**: 57–76
- Chaitin GJ (1974) Information-theoretic limitations of formal systems. *JACM* **21**: 403–424
- Cook SA (1971) The complexity of theorem proving procedures. In *Proceedings of the Third*

- Annual ACM Symposium on the Theory of Computing* pp 151–158. New York, NY, USA: Association for Computing Machinery
- Cronin L *et al* (2006) The imitation game—a computational chemical approach to recognizing life. *Nat Biotech* **24**: 1203–1206
- Danchin A (2009) Bacteria as computers making computers. *FEMS Microbiol Rev* **33**: 3–26
- di Ventura B, Lemerle C, Michalodimitrakis K, Serrano L (2006) From *in vivo* to *in silico* biology and back. *Nature* **443**: 527–533
- Endy D (2005) Foundations for engineering biology. *Nature* **438**: 449–453
- Fisher J, Henzinger TA (2007) Executable cell biology. *Nat Biotech* **27**: 1239–1249
- Gibson DG *et al* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215–1220
- Gil R, Silva FJ, Peretó J, Moya A (2004) Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* **68**: 518–537
- Gödel K (1931) Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik* **38**: 173–198
- Goss PJ, Peccoud J (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc Natl Acad Sci USA* **95**: 6750–6755
- Harel D (2005) A Turing-like test for biological modeling. *Nat Biotech* **23**: 495–496
- Keller EF (2002) *Making Sense of Life. Explaining Biological Development with Models, Metaphors, and Machines*. Cambridge, MA, USA: Harvard University Press
- Lartigue C, Glass JL, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. *Science* **317**: 632–638
- Luisi PL, Ferri F, Stano P (2006) Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften* **93**: 1–13
- Mansy SS, Schrum JP, Krishnamurthy M, Tobé S, Treco DA, Szostak JW (2008) Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**: 122–125
- Mayr E (2002) *The Autonomy of Biology*. Walter Arndt Lecture. http://www.biologie.uni-hamburg.de/b-online/e01_2/autonomy.htm
- Morange M (2005) *Les Secrets du Vivant. Contre la Pensée Unique en Biologie*. Paris, France: Éd. La Découverte
- Moya A, Gil R, Latorre A, Peretó J, Garcillán-Barcia MP, de la Cruz F (2009) Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol Rev* **33**: 225–235
- Penrose R (1989) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford, UK: Oxford University Press
- Pereira de Souza T, Stano P, Luisi PL (2009) The minimal size of liposome-based model cells brings about a remarkably enhanced entrapment and protein synthesis. *Chem Biochem* **10**: 1056–1063
- Peretó J, Català J (2007) The renaissance of synthetic biology. *Biol Theor* **2**: 128–130
- Rasmussen S, Bedau MA, Chen L, Deamer D, Kraukauer DC, Packard NH, Stadler PF (2008) *Protocells: Bridging Nonliving and Living Matter*. Cambridge, MA, USA: MIT Press

Regev A, Shapiro E (2002) Cells as computation. *Nature* **419**: 343

Romero-Campero FJ, Pérez-Jiménez MJ (2008a) A model of the quorum sensing system in *Vibrio fischeri* using P systems. *Artif Life* **14**: 95–109

Romero-Campero FJ, Pérez-Jiménez MJ (2008b) Modelling gene expression control using P systems: the Lac operon, a case study. *Biosystems* **91**: 438–457

Sadot A, Fisher J, Barak D, Admanit Y, Stern MJ, Hubbard EJ, Harel D (2008) Towards verified biological models. *IEEE/ACM Trans Comput Biol Bioinform* **5**: 223–234

Serrano L (2007) Synthetic biology: promises and challenges. *Mol Syst Biol* **3**: 158

Szallasi Z, Stelling J, Periwal V (2006) *System Modeling in Cellular Biology*. Cambridge, MA, USA: MIT Press

Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. *Nature* **409**: 387–390

Turing A (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proc Lond Math Soc Series 2* **42**: 230–265

Left to right: Andrés Moya is at the Departament de Genètica and Institut Cavanilles de Biodiversitat i Biologia Evolutiva, University of València, Centro Superior de Investigación en Salud Pública (CSISP) and CIBER de Epidemiología y Salud Pública (CIBEResp), Spain.

E-mail: andres.moya@uv.es

Natalio Krasnogor is at the School of Computer Science, University of Nottingham, UK.

E-mail: natalio.krasnogor@nottingham.ac.uk

Juli Peretó is at the Departament de Bioquímica i Biologia Molecular and Institut Cavanilles de Biodiversitat i Biologia Evolutiva, University of València and CIBER de Epidemiología y Salud Pública (CIBEResp), Spain.

E-mail: juli.pereto@uv.es

Amparo Latorre is at the Departament de Genètica and Institut Cavanilles de Biodiversitat i Biologia Evolutiva, University of València, Centro Superior de Investigación en Salud Pública (CSISP), and CIBER de Epidemiología y Salud Pública (CIBEResp), Spain.

E-mail: amparo.latorre@uv.es

doi:10.1038/embor.2009.120