

***IN SILICO* BIOLOGY**

CHRISANTHA T. FERNANDO

*Department of Informatics, University of Sussex
Falmer, Brighton, BN1 9RH, United Kingdom*

*Mathematical Biology, National Institute for Medical Research
Mill Hill, London, NW7 1AA, United Kingdom*

*Mathematical Collegium Budapest (Institute for Advanced Study),
Szentháromság u. 2, H-1014, Budapest, Hungary*

RICHARD A. GOLDSTEIN

*Mathematical Biology, National Institute for Medical Research
Mill Hill, London, NW7 1AA, United Kingdom*

PHIL HUSBANDS

*Department of Informatics, University of Sussex
Falmer, Brighton, BN1 9RH, United Kingdom*

DOV J. STEKEL

*School of Biosciences, University of Nottingham, Sutton Bonington Campus
Nottingham, LE12 5RD, United Kingdom*

Rather than studying existent living systems, we can increasingly produce computer models that capture the salient aspects of life. This provides us with unprecedented opportunities to examine, manipulate, and explore biological phenomena, allowing us to investigate some of the deepest issues in biology.

1. Introduction

It's life, Jim, but not as we know it [1].

The organisms that we study today are the outcome of a specific, historically contingent process of billions of years of evolution. Over its history, biology has mostly involved the description and classification of observed life forms. We have made great progress in that time in expanding the scale of our observations, looking at a much wider range of organisms using increasingly precise and detailed measurement techniques. Especially with the increasingly data rich experimental techniques of genetics and molecular biology, we have coupled these investigations with mathematical modelling and computer simulations, developing our understanding of increasingly large and complex biological systems, elucidating the functions of these systems, and proposing hypotheses about their evolution.

Despite these advances, there are severe limitations on our current techniques. Firstly, there is much information about living systems that is either practically or inherently difficult or impossible to obtain. We cannot measure systems with arbitrary precision. Many measurement techniques either look at bulk properties of large ensembles of components, or alternatively, measure the distinct properties of a small number of these components, rather than allowing us to look in a comprehensive way at distributions (including outliers). Measurement methods explore a limited range of timescales, and cannot be arbitrarily extended to very slow or very fast processes. It is generally impossible to study a system without the examination itself perturbing the object of study. Studying living systems are particularly difficult, as the requirements of life are incompatible with many investigative techniques.

Conversely, there are many manipulations that we might wish to make but cannot. We cannot arbitrarily change environmental conditions. We cannot change the properties of the physical components that make up living systems.

2

We cannot delete species from food webs. We cannot investigate the role of stochastic fluctuations by removing them from the system.

Evolutionary investigations offer their own difficulties. We cannot directly explore the process that gave rise to current organisms, running evolution under controlled and monitored conditions on the time scales for which observed organisms have evolved. We cannot know the relative importance of ‘Chance and Necessity’ [2] in explaining observed biology. We cannot explore avenues that evolution might have taken under different circumstances; we cannot ‘play the tape twice’ [3].

Finally, it is difficult to deduce general principles from the single example represented by the current set of evolutionarily related organisms. We cannot experimentally determine whether universal features found in all living organisms represent absolute constraints, common responses to a common situation, or a randomly chosen alternative maintained through a common genetic descent. We cannot address questions about what would have happened if it had arisen under different circumstances.

Rather than analysing specific living systems, we can try to create artificial systems that have properties characteristic of natural living systems [4]. Experimentalists have had some success investigating fundamental issues in evolutionary biology through, for instance, directed *in vivo* evolution [5, 6]. Progress has also been made by either building up cells from non-living components [7, 8] or stripping existent cells of all but the most essential ingredients [9]. The experimental challenges of this work are daunting, and there remain many limitations to the possibilities that can be explored.

An alternative is *in silico* biology, creating computational simulations of systems that embody important properties of biological systems, such as homeostasis, reproduction, evolution, etc. This approach allows us to explore, investigate, and manipulate the workings of such systems without being restricted to the particular examples nature has made available to us. We can explore the processes of Darwinian evolution by studying digital ‘organisms’ on true evolutionary timescales not possible in the laboratory [10], including their origin [11-15], evolution [16-18], development [19, 20], behaviour [21], and interactions [22, 23]. The aim of these simulations is often the discovery of general or fundamental principles and organization of life, rather than purely the elucidation of the workings of a specific example.

In silico biology represents a movement of biology from an analytical to a synthetic framework as we create artificial systems that exhibit biological behaviour. This shift has three consequences. Firstly, we can greatly expand our subjects of investigation beyond those that are naturally available. In this way, we can go beyond the idiosyncratic examples provided by nature, allowing us to generate a set that is both more varied and more systematic. Secondly, it is by attempting to create artificial systems that embody many of the phenomena of life that we most rigorously test our understanding of biology, investigate what is sufficient, what is necessary, what is understood, what is missing. Finally, in addition to providing a deeper theoretical foundation for biology, these activities can provide important tools for computer science as evidenced by the success of various evolutionary computational methods.

2. Work in *in silico* biology

The first investigations of *in silico* biology were the introduction of automata by John von Neumann in the late 40s. . Since then, this field has grown to include many topics, including

- Evolution of diversity and complexity. Evolution by natural selection has produced a biosphere with more than 10 million species and organisms ranging in size by 10 orders of magnitude. Yet although evolution underpins all of biology, we do not yet understand how and why the observed diversity and complexity arises. A major contribution of *in silico* Biology is the capacity to explore our understanding of evolution by testing what factors are necessary to observe the emergence of complex life forms. Specifically, we are interested in understanding how it is possible for complex traits to emerge in an open-ended manner [24], or indeed why complexity emerges at all [25]. Particularly important is garnering an understanding of how and why the major transitions in evolution, such as the evolution of multicellularity or of sexual reproduction, have arisen [26]. Because these cannot be repeated, *in silico* techniques are particularly valuable.

- The evolution of evolvability: Selective pressures exist at different levels, including genetic, individual, kinship group, lineage and species group. *In silico* biology can help us to understand the relationship between these levels of selection [27]. Of particular interest is understanding whether the capacity for evolution is itself a property that evolves [28-30], and, if so, under what circumstances [17].
- Biological Networks: What governs the properties of networks of dynamical elements, such as those that occur in biochemical and gene-regulatory networks [31, 32]? How do they evolve [33]? Why are some network architectures much more common than would be expected at random [34, 35]? How do we understand the robustness and modularity observed in such networks [36]? How does gene-expression produce morphology [37]? How can we analyse biological networks [38]?
- Co-evolution: How do we understand the relationship between hosts and pathogens [39]? How has this relationship determined our evolutionary process? How does this affect our expectations about newly emergent diseases?
- Learning and Evolution: How does learning and the ability to learn affect the evolutionary process [40-44]? How are collections of relatively simple organisms (such as an ant colony or wasp hive) able to exhibit intelligent behaviour far in advance of that of the individuals? Can bacteria undertake associative learning? [45-47].
- The Evolution of Communication and Signalling: How does communication arise [48, 49]? What are the necessary conditions for its emergence [50-52]?
- The Origin of Life: What are the most basic features required for something to be 'living' [14]? How did life arise [26]? How did nucleotides arise [13]? What was the function of the first replicase ribozyme [53]?
- Emergence: One theme throughout in *in silico* biology is the important process by which complex behaviour can result through the interactions of a large number of relatively simple components. How can we understand emergence [54]?

3. Things as they are, things as they are not

You see things; and you say 'Why?' But I dream things that never were; and I say 'Why not?' [55]

The question that the Serpent asks Eve in Shaw's play makes a distinction between 'Why' and 'Why not' questions. But in reality, these questions are often the same. If we wish to understand why things are the way they are, we need to look at alternatives and consider the reasons that these alternatives did not occur. Unfortunately, the world has only provided us with a limited set of highly related living organisms. *In silico* biology can provide further alternatives, giving us increased understanding regarding the conditions under which such alternatives might have arisen, and why the currently observed organisms arose instead.

References

1. *Star Trek*, G. Roddenberry, creator.
2. J. Monod, *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*, Vintage, London (1972).
3. W. Fontana and L. W. Buss, *Proc. Natl. Acad. Sci. USA*, **91**, 757 (1994).
4. R. E. Lenski, C. Ofria, T. C. Collier and C. Adami, *Nature* **400**, 661 (1999).
5. M. Travisano, J. A. Mongold, A. F. Bennett and R. E. Lenski, *Science*, **267**, 87 (1995).
6. J. A. G. M. De Visser and D. E. Rozen, *J. Evol. Biol.*, **18**, 779 (2005).
7. C. Fernando, M. Santos and E. Szathmáry, *Top. Curr. Chem.*, **259**, 167 (2005).
8. P. L. Luisi, *The Emergence of Life: from Chemical Origin to Synthetic Biology*, Cambridge University Press, Cambridge(2006).
9. A. C. Forster and G. M. Church, *Mol. Syst. Biol.*, **2** (2006).
10. E. Szathmáry, *Philos. Trans. R. Soc. London. B. Biol. Sci.*, **361**, 1761 (2006).
11. E. Szathmáry and J. Maynard Smith, *J. Theor. Biol.*, **187**, 555 (1997).
12. C. Fernando and J. Rowe, *J. Theor. Biol.* **247**, 152 (2007).
13. C. Fernando and J. Rowe, *Biosys.*, **91**, 355 (2008).
14. T. Gánti, *The Principles of Life*, Oxford University Press, Oxford, UK (2003).

15. R. Lathe, *Icarus*, **168**, 18 (2003).
16. C. Adami, *Introduction to Artificial Life*, Springer (1999).
17. M. Lynch, *Proc. Natl. Acad. Sci. USA* **104**, 8597 (2007).
18. J. P. Crutchfield and P. Schuster, *Evolutionary Dynamics*, OUP (2002).
19. K. Stanley and R. Miikkulainen, *Artificial Life*, **9**, 93 (2003).
20. I. Salazar-Ciudad and J. Jernvall, *Evol. Devel.* **6**, 6 (2004).
21. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall, (2006).
22. J. Hofbauer and K. Sigmund, *Evolutionary games and population dynamics*, Cambridge University Press, Cambridge (1998).
23. R. Axelrod, *The Evolution of Cooperation*, Basic Books (1985).
24. M. A. Bedau, *Artificial Life*, **4**, 125 (1998).
25. C. Adami, C. Ofria and T. C. Collier, *Proc Natl Acad Sci U S A.*, **97**, 4463 (2000).
26. J. Maynard Smith and E. Szathmáry, *The Major Transitions in Evolution*, Oxford University Press, Oxford (1995).
27. S. Okasha, *Evolution and the levels of selection*, Oxford University Press, Oxford (2006).
28. M. Pigliucci, *Nat. Rev. Genet.*, **9**, 75 (2008).
29. A. G. Jones, S. J. Arnold and R. Bürger, *Evolution*, **61**, 727 (2007).
30. W. Zhu and S. Freeland, *J. Theor. Biol.*, **239**, 63 (2006).
31. R. Thomas, *Int. J. Dev. Biol.* , **42**, 479 (1998).
32. T. Lenser, T. Hinze, B. Ibrahim and P. Dittrich, in *Proceedings EvoBio* Vol. LNCS 4447, Marchiori, E., Moore, J. H., and Rajapakse, J. C., Eds., Springer Verlag, Valencia, pp. 132 (2007).
33. M. M. Babu, S. A. Teichmann and L. Aravind, *J Mol Biol*, **358**, 614 (2006).
34. N. Kashtan and U. Alon, *Proc Natl Acad Sci USA*, **102**, 13773 (2005).
35. O. Sporns and R. Kotter, *PLoS Biology*, **2**, e369 (2004).
36. A. Wagner, *Robustness and Evolvability in Living Systems*, Princeton University Press, Princeton (2007).
37. G. B. Muller, *Nat. Rev. Genet.*, **8**, 943 (2007).
38. J. Bongard and H. Lipson, *Proc Natl Acad Sci U S A.*, **104**, 9943 (2007).
39. M. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*, Harvard University Press, Cambridge (2006).
40. P. Adams, *J Theor Biol*, **195**, 419 (1998).
41. M. J. Baldwin, *Psychol. Rev.*, **16**, 207 (1909).
42. M. J. Baldwin, *Am. Natur.*, **30**, 441 (1896).
43. M. J. Baldwin, *Psycholog. Rev.*, **5**, 4 (1898).
44. G. E. Hinton and S. J. Nowlan, in *Adaptive Individuals in Evolving Populations. Models and Algorithms*, Addison-Wesley Longman, 447 (1987).
45. I. Tagkopoulos, Y.-C. Liu and S. Tavazoie, *Science*, **320**, 1313 (2008).
46. C. Fernando, A. M. L. Liekens, L. E. H. Bingle, C. Beck, T. Lenser, D. J. Stekel and J. E. Rowe, *J. Roy. Soc. Interface* **6**:463 (2009).
47. N. Gandhi, G. Ashkenasy and E. Tannenbaum, *J. Theor. Biol.*, **249**, 58 (2007).
48. H. Brighton, in *Theoretical and Applied Linguistics*, Vol. PhD, The University of Edinburgh, Edinburgh (2003).
49. L. Steels, R. van Trijp and P. Wellens, in *ECAL07*, Almeida e Costa, F. a. R., L.M. and Costa, E. and Harvey, I, Ed., Springer Verlag. , pp. 425 (2007).
50. L. Steels, *Trends in Cognitive Sciences*, **10**, 347 (2006).
51. M. Oliphant and J. Batali, *The newsletter of the Center for Research in Language*, **11**, 1 (1997).
52. E. Szathmáry and S. Szamado, *Nature*, **456**, 40 (2008).
53. C. Fernando, G. Von Kiedrowski and E. Szathmáry, *J. Mol. Evol.*, **64**, 572 (2007).
54. S. McGregor and C. Fernando, *J. Artif. Life*, **11**, 459 (2005).
55. G. B. Shaw, *Back to Methuselah*, (1921).