A partnership between biology and engineering

Roger Brent

This article explores the potentially beneficial outcomes of a partnership between systems biology and synthetic biology. This assessment is a challenge due to the vague definition and unrealistic claims made for systems biology, as well as by the lack of an explicitly stated distinction between synthetic biology and the engineering of biological systems practiced since the development of recombinant DNA. Here, I suggest that one might be able to add meaning to the concept of systems biology by remembering older conceptions of experimental systems. In biology, the original word used for the study of system function is physiology. It may be possible in the near term to understand the quantitative physiology of certain intracellular systems. I then try to determine the distinguishing attributes of synthetic biology. Any body of theory and experimental capability that enables quantitative prediction of a system's behavior will be applicable to synthetic biology in that it will enable prediction of the behavior of human-designed biological artifacts before those are instantiated in DNA code. If the practitioners are honest with one another about the limits of their abilities, this intersection of science and engineering can spur the development of both.

Engineered biological systems and synthetic biology

Among scientists in the 1960s, the early successes of molecular biology gave rise to widespread belief that new synthetic technologies would develop[1]. By the next decade molecular biology began to impact human synthetic capability in earnest through the development of recombinant DNA methods[2,3]. Before the 1970s, human biological synthetic ability had been increasing gradually since the Neolithic, with a noticeable uptick around the Enlightenment and another in the early 20th century with the development of transmission genetics. But in the 1970s, the rate of increase went high and stayed high. A bare ten years after the first publications, one could purchase bottles of human insulin, synthesized in bacteria, in any pharmacy in the developed capitalist world.

The origin of what is now called synthetic biology is more recent[4]. Although a thoughtful observer has recently called this a scientific discipline, akin to synthetic biology[5], here, I will use this term to describe efforts to design biological systems to perform a given function, verify that they will have that function before one builds them, instantiate them in DNA code and use the system to accomplish the function. This nexus of effort is obviously a branch of engineering rather than science. But to understand how what is sometimes called systems biology might support the development of this engineering, one needs to introduce the science.

Imprecise definitions and inadequate methods

Although common use of the term 'systems biology' is fairly new, the concept is not. The current vogue for the term can be understood as a dialectical swerve, a consequence of the success of the molecular biological program articulated by Warren Weaver in the 1930s: to understand life, understand the molecules that make it up[6]. By the 1990s, and led by the genome projects, various kinds of high-throughput data collection projects were coming online, and their output was of course information about molecular entities (coding sequences, mRNAs, enzymatic functions, post-translational modifications and contacts with other proteins). If these activities achieved their goals, biologists would possess descriptions of molecules, times and places of RNA and protein synthesis, simple biochemical functions and their connections[7]. The word for descriptions of parts and their connections is anatomy. The work for how those parts work together to provide function is physiology. The 1990s was a festival of information about molecules. We were due for a move back to physiology.

Systems biology is usually now taken to mean that in order to understand the behavior of a biological ensemble, one needs to study the whole, rather than its isolated parts. At that level, the idea is both obvious and devoid of deeper explanatory power[8]. The term evokes 'systems neuroscience,' which arose in opposition to 'molecular neuroscience,' a meme from the 1980s when recombinant DNA–powered molecular methods began to impact neurobiology departments. In neuroscience, neither usage is profound, because understanding of the brain obviously needs to be grounded in the components that make it up (the molecular and cellular bits), but one obviously then needs to understand how molecules and cells work together to give rise to functions (the systems bit). In fact, neurobiology started with systems approaches long before individual molecules had been identified, and this research tradition stretches from the study of sensory systems pioneered by von Helmholtz, through contemporary studies of phenomena that depend on ensembles of small numbers of neurons, such as long-term potentiation. These successful examples of systems neuroscience have in common a more sophisticated definition of the word system, to which I return below.

In molecular, cellular and developmental biology, proponents of systems biology sometimes promised, explicitly or not, that replacing the trickle of data about the molecules

Roger Brent is at The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, California 94704, USA.
e-mail: brent@molsci.org

of life with a torrent would enable wholesale classification and analysis of the data (which it did), but also that some foreseeable analysis of the catalog could not just supplement but obviate the need for directed experimentation (which it did not).

Some of the criticisms directed against this picture of systems biology are definitional. At the margins, a great deal of science that uses of high-throughput information is indistinguishable from other research. For example, suppose that analysis of gene expression data suggests that a set of genes might be controlled by a positive regulator. An experimenter might scan a list of genomic open reading frames to identify possible DNA-binding transcription activators, make cells in which these were knocked out, survey expressed mRNAs and find that one of the mutated candidate transcription activators was needed for expression of that set of genes. This experimenter might call the line of experimentation systems biology, another might not, but the name one gives the type of experimentation that led to the conclusion seems of little consequence.

But some of the criticisms one can direct at the claimed capabilities go beyond terminology. Of these, two seem particularly important. One is the idea that high-throughput data can be sufficient by themselves to learn things about a cell, organism or pathway that are not already known. It is certainly true that surveys bring to one's attention facts that would not otherwise have been found. For example, early large scale surveys found that genes induced during fibroblast growth were involved in wound healing[9], an observation that made great sense after the fact but which was not widely appreciated beforehand, and a great deal of recent work is enabling new delineations of types of cancer cells and their response to therapies. But in these cases, as in the thought experiment about activator proteins above, the new data is only useful taken together with prior knowledge of how things work. The other idea is closely entangled: that there might be experimental perturbations that can be applied systematically and paired with existing high-throughput data collection methods to yield systematic insight. This vision is not silly. For example, it has been possible to use data on changes in yeast mRNA expression in response to essentially arbitrary chemical and genetic perturbations to identify candidate genes that might be expressed in specific cell lineages[10]. But, even in the rare successful cases, the vision is incomplete—at the moment, the new hypotheses need to be tested by *ad hoc* experimentation. Moreover, this vision is also premature, because, no matter what perturbations one can generate systematically, the types of data one can now generate systematically to observe their consequences are not adequate to determine either the architectures or quantitative behaviors of the regulatory pathways of most interest to contemporary biologists (see below).

Productive systems

One way to get beyond current inadequacies of the systems biology concept may be to 'rebrand' it by recalling an older and more precise conception of system. Consider this definition from the computer scientist Ben Kuipers in 1994 (ref. 11): "By a system, we simply mean some subset of the entire world whose behavior, and whose interaction with the rest of the world, we believe can be sensibly described. The set of all clock pendulums in New England is undeniably a subset of the world.... but is not a good example of a system." In the hands of an experimentalist like Galileo, the ensemble of components (weight, arm, hinge) that comprises a single clock pendulum becomes a system, in fact, a good system. Certain attributes make it so. First, the experimenter can draw a boundary around it so that, to a first approximation, variable influences from outside that boundary can be ignored. The boundary is drawn heuristically; the experimenter can redraw it at will. Second, the experimenter can conduct defined perturbations inside the boundaries (for example, lengthening the arm of the pendulum) that lead to changes in system behavior. Third, the experimenter can, by reasoning, generate sensible explanations for the changed behavior. Fourth, the 'goodness' of the system is defined after the fact, by assessing how much the investigator learned from it. Fifth, in many cases an operational means of assessing how much the investigator learned is to evaluate how well the understanding enables prediction of changed system behavior in response to defined perturbations. In biology, the practice of investigating function by delineating a subset of the world and perturbing components within the boundary is old indeed. It goes back at least to William Harvey and the idea of a circulatory system, and in genetics and molecular biology stretches in an unbroken line from T.H. Morgan through the work of Jacob and Monod (who defined and made good use of lactose metabolism and λ lysogeny as experimental systems) to the present day[12].

In biology, some subsets of the world that we would like to understand are now beyond our approach, but will not be so in the future. For example, imagine that we could generate an inventory of the expression and subcellular and extracelluar localization of all mRNAs and proteins expressed by a genome in all the cells of an organism over developmental time. At the moment, this inventory of expressed molecules would not constitute a system in the Kuipers sense because its detailed behavior could not be sensibly described. However, this inventory differs from the New England pendulums because its temporal evolution is at least partially determined by its genomic source code. Today, we can sequence the code. One day we will be able to sensibly describe its workings over developmental time. On that day, the inventoried expression and localization patterns will merit the term system. And we will call the conceptual framework that enabled the sensible description of those patterns a theory.

Prediction as understanding

The reason for paying attention to theory is that it—aside from the ability to supply parts—may provide a path by which biology, or systems biology, will contribute to synthetic biology.

To illustrate this point, let us suppose that there is some combination of theory and experimental data for a biological system that would enable prediction of its future quantitative behavior given knowledge of its present state. This stance admits that some aspects of biological behavior are not determined, but acknowledges that many aspects are highly determined. The developmental trajectory of entire metazoan organisms in time and space is determined to a large extent by the elaboration of instructions stored in their genomes. On shorter time scales, the behavior of a cell 60 seconds after a perturbation is dependent on the makeup of a constellation of molecules already present at the time of perturbation. But attaining quantitative insight into this behavior, even for simple systems on short time scales inside single cells, will be extremely difficult. One reason is that the only current theoretical frameworks that seem to allow the desired predictive power require measuring numbers of molecules and the rates at which they participate in defined reactions. Another is that, methods to measure these quantities do not now exist, but need to be developed. And, even if one can achieve this understanding, one cannot now guess its pure scientific worth.

However, when one assesses the possible value of quantitative prediction to applications of biological knowledge, the picture changes. Predictive quantitative frameworks that described even quite small systems, for example, intracellular signal transduction

systems, would find use in medicine, both to guide drug discovery efforts, and to guide therapy by suggesting points, perhaps in individual cell types of individual patients, for which small, perhaps multiple simultaneous interventions might have great therapeutic value. Another area for application of this understanding would be in synthetic biology.

Biological engineering, synthetic biology and reinvented wheels

The comparison between biological systems and human engineered artifacts became compelling during the late 1990s (see, for example, refs. 13,14,15,16). Today, numerous researchers want to design and build biological systems, and call this work 'synthetic biology'[4,17,18,19,20,21,22]. These researchers distinguish their work from an older biological engineering canon, which encompasses fermentation and process engineering, and also biomedical engineering (prosthetic limbs, laser catheters guided through arteries, cochlear implants). They also distinguish it from a second, newer and economically important, recombinant DNA canon, which encompasses engineered organisms that produce proteins or simple chemicals, plants engineered that make pesticides or fix nitrogen, phage vectors designed with attributes relevant to gene therapy. Instead, the synthetic biologists have defined as their goal the design and construction of systems that exhibit complex dynamical behavior, logical behavior, the ability to exist in a number of states or the ability to execute small numbers of programmed steps (for example, in complex chemical synthesis).

Some of this engineering work has a distinct feeling of a new group of investigators revisiting a body of existing knowledge, so that the conclusions arising from it are difficult to distinguish from previous work. Some of the ideas are quite old: the first genetic oscillatory system to be defined (in 1970), cI and cro[23] (Eisen *et al.*, 1970) came from then-cutting-edge biology, the study of phage λ. Among themselves, these researchers referred to this subsystem as a 'circuit,' and named it after a then-current digital electronic artifact, a 'flip flop,' a usage that seems even more prescient when one remembers that these circuits were then soldered together from discrete transistors and that this English term was being used by scientists speaking French. Similarly, molecular schemes, some involving quite complex engineering that take advantage of DNA-binding proteins and regulatory sites to control gene regulation[24–27], have been around for many years. It is by no means clear at this point that the attempts by synthetic biologists to characterize the behavior of their systems has resulted in insights beyond those gained by the classical biologists who studied the systems they constructed, although it may well happen in the future. In any case, the worth of these latter-day engineering efforts will need to be established in engineering terms by development of new functionalities that see wide use.

Similarly, the criteria that distinguish the synthetic biology canon from the second canon of recombinant DNA and genetic engineering are at this point not well defined. One seeming distinction is that synthetic biologists state their desire to assemble complex living systems from well-defined parts, rather than by making simpler interventions that suborn existing systems. For example, the effort to use combinations of regulatory proteins and sites to engineer a cell that can add two numbers together might be considered synthetic biology, whereas the effort to modify the glycosylation machinery of maize to enable production of a human antibody might not. A second distinction is that the people performing the engineering work call themselves engineers. So workers constructing the cell that adds two numbers might call themselves engineers practicing synthetic biology, workers making the maize that makes a properly glycosylated antibody might call themselves scientists and workers modifying a gene therapy vector for use in a clinical trial may call themselves physicians. A third distinction is that synthetic biologists positively emphasize a number of doctrinal points from other branches of engineering, including the use of well-behaved interchangeable parts, the separation of design from fabrication, the taming of complexity and mitigation of unexpected consequences by disciplining the engineering to require rigidly defined levels of abstraction (part, device, system) with protocols for communication among those levels, and of course the willingness to use parts and systems, such as aminoacyl tRNA synthetases that work with nonstandard amino acids, that are not found in the natural world[28].

Synthetic biologists have defined as their goal the design and construction of systems that exhibit complex dynamical behavior.

For the purpose of this article, one of these doctrinal points is paramount. That is an emphasis on design and testing via simulation before fabrication. Synthetic biologists wish to examine the performance of the systems they design before instantiating them in DNA code. Let us call this a longing for 'design-based engineering of biological systems.' Whatever synthetic biology means now or is to mean in 10 years (and this we can leave up to the practitioners), it is here that the intersection between it and systems biology, defined or rebranded as the quantitative biology of function, the physiology, of ensembles of defined parts, seems to lie.

Partnership and consequences

For biologists, this intersection has at least two strongly positive consequences. The first is the infusion of engineering ideas and mindsets into the quest to understand the quantitative biology of function. That scientific objective will not be accomplished without the contributions of biologists, experimental physicists, chemists, applied mathematicians and computer programmers, and because of that need, the cross-cultural fertilization at research sites working on this problem is tremendous. Adding engineering to the mix only increases the complexity of the cultural collision and the potential for useful ferment. For example, the hacker culture of software engineering emphasizes elegance of execution and playfulness, qualities surely to be prized among research scientists. Similarly, software engineers often consider open source development paths for technology[29,30], and, if these concepts can be stretched further, they may offer alternative ways to organize human efforts to design and construct complex biological systems. The second is that some scientific and engineering activities are essentially identical but have different names. For example, consider a computer program that runs simulations of the time-dependent behavior of a biological system. A biologist might consider this a quantitative predictive model. A synthetic biologist might look at the same piece of code and call it a design tool. Here, progress toward one goal is essentially equivalent to progress toward the other.

For engineers, or synthetic biologists, the intersection with biology may have also positive consequences. One reason is that at this point, biologists usually know a great deal more about how to work with biological systems. If an engineer wants to effect a design goal, her ability to execute it is likely to benefit from advice from a scientist. Another reason is that engineering has produced concepts—wheels and axles, governors on steam engines,

impedance in electrical circuits, switchboards in telephone exchanges, synchronous operations in digital processors—all of which were designed by sentient beings instead of being thrown up by evolution. Engineers, naturally, now seek to test the relevance of these concepts to biology. In any given case, an analogy might be fruitful or might not. It is usually the biologists who now attempt to evaluate the utility of the engineering analogies, and their feedback is at least of some value to engineers.

The above gives examples of the possible benefits of cross-fertilization between science and engineering. For these benefits to be realized, the proper stance from both sides is probably one of skepticism tempered by respect. There is now no systems biology that engineers can use to predict the behavior of the systems they design. There is now no universally applicable set of concepts from the behavior of mechanical or electrical systems that biologists can immediately apply to gain deeper understanding of the cells and organisms they study. Cells and organisms will need to be understood in their own terms. Many of those terms will come from science. But some will come, as they have already, from future concepts in engineering.

The long-term reason to pay attention to this intersection is that the main way in which increases in biological understanding affect human affairs is by enabling increases in human capability. A number of people[7,31] have pointed out that biological capability, the ability to manipulate and build living things, is likely to become as important to the economy of this century as the ability to manipulate the flow of electrons and bits and build circuits was to the last. A predictive biological understanding—or, if we are still using the term, a systems biology that is more than a collection of slogans—is the natural marriage partner for a synthetic biological capability. This partnership would enable vastly more sophisticated therapies, and more precise engineering of crop plants. But its impact would not be confined to the existing healthcare and agricultural industries. Devices that counted cell divisions or that bar-coded different metaphyte and metazoan lineages during differentiation could move from tools of discovery to tools to monitor the real-time condition of economically important plants and animals. Genetic circuits installed as transient tattoos by ballistic DNA transformation of the skin could monitor exposure to radiation and environmental toxins. Design-based biological engineering could result in the ability to program and execute multistep syntheses that could call on chemistries evolution never generated, and could then lead to the ability to design collections of cells and organisms, tailored ecologies, that execute combinations of chemical steps (synthesis) and mechanical steps (fabrication). Such capabilities could obviously affect the generation of energy, materiel, roads, housing and clothing, and fabrication of more complex artifacts such as vehicles and computers. As importantly, this partnership will eventually enable the wholesale design-based, predictive modification of organisms important to us, including, if we so choose, our own species.

ACKNOWLEDGMENTS

I am grateful to Drew Endy, Tom Knight, Gerald Sussman, Rob Carlson, Larry Lok and Harvey Eisen for useful discussions over many years, and to Paul Rabinow, Orna Resnekov, Myron Williams and the anonymous reviewers for criticisms that greatly improved this manuscript. Work is supported under Alpha Project at the Center for Genomic Experimentation and Computation, an NIH Center of Excellence in Genomic Science, supported by grant P50 HG02370 to R.B. from the National Human Genome Research Institute.

1. "It is conceivable that by 1984 we shall produce our food exploiting … the synthesis of proteins in cell free systems. Eventually we should be able to manufacture satisfactory foodstuffs in great chemical plants, where masses of ribosomes would be supplied with synthetic amino acids and long-lived messenger-RNAs, with energy-yielding phosphates supplied by irradiating chloroplasts with laser-tuned light of the most effective wavelength." Waddington, C.E.F. "Science and wisdom" in *The World in 1984, the Complete New Scientist Series*, vol. 2 (ed. Nigel Calder) (Penguin Books, Harmonsdworth, Middlesex, 1965).
2. Cohen, S.N., Chang, A.C.Y., Boyer, H. & Helling, R.B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl. Acad. Sci. USA* **70**, 3240–3244 (1973).
3. Rodgers, M. The Pandora's box congress. *Rolling Stone* **189**, 37–77 (1975).
4. Gibbs, W.W. Synthetic life. *Sci. Am.* **290**, 75–81 (2004).
5. Benner, S.A. Synthetic biology: act natural. *Nature* **421**, 118 (2003).
6. Kay, L. The Molecular Vision of Life, Caltech, The Rockefeller Foundation, and the Rise of the New Biology (Oxford University Press, New York, 1993).
7. Brent, R. Genomic biology. *Cell* **100**, 169–183 (2000).
8. "Systems biology is a loosely defined term, but the main idea is that biology is an information science, with genes a sort of digital code. Moreover, while much of molecular biology has involved studying a single gene or protein in depth, systems biology looks at the bigger picture, how all the genes and proteins interact." Pollack, A. Scientist at work: Leroy Hood. *New York Times* April 17 (2001), p. F3
9. Iyer, V.R. *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83–87 (1999).
10. Colman-Lerner, A., Chin, T. & Brent, R. Yeast Cbk1 and Mob2 activate daughter-specific genetic programs to induce asymmetric cell fates. *Cell* **107**, 739–750 (2001).
11. Kuipers, B. Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge (MIT Press, Cambridge, Massachusetts, 1994).
12. "It may be interesting to realize that the above connotations of system (thing experimented on; heuristically defined, semi-permeable, and shifting boundary between it and rest of world, etc.) are not intuitive to some computer scientists. These researchers of this "science of the artificial" (Simon, 1981) never faced the need to isolate approximately and heuristically their objects of study from the larger world, because they could do so absolutely, simply by writing the proper code." Simon, H.R. *The Sciences of the Artificial*, edn. 1 (MIT Press, Cambridge, Massachusetts, 1981).
13. McAdams, H.H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* **94**, 814–819 (1997).
14. McAdams, H.H. & Arkin, A. Towards a circuit engineering discipline. *Curr. Biol.* **20**, R318–R320 (2000).
15. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
16. Elowitz, M. & Leibler, S. A synthetic oscillatory network of transcription regulators. *Nature* **404**, 335–338 (2000).
17. Ferber, D. Microbes made to order. *Science* **303**, 158–161 (2004).
18. Yokobayashi, Y., Weiss, R. & Arnold, F.H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA* **99**, 16587–16591 (2002).
19. Atkinson, M.R., Savageau, M.A., Myers, J.T. & Ninfa, A.J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
20. Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T. & Weiss, R. Spatiotemporal control of gene expression with pulse generating networks. *Proc. Natl. Acad. Sci. USA* **101**, 6355–6360 (2004).
21. Kobayashi, H. *et al.* Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA* **101**, 8414–8419 (2004).
22. Kramer, B.P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004).
23. Eisen, H., Brachet, P., Pereira da Silva, L. & Jacob, F. Regulation of repressor expression in λ. *Proc. Natl. Acad. Sci. USA* **66**, 855–862 (1970).
24. Maurer, R., Meyer, B. & Ptashne, M.J. Gene regulation at the right operator (OR) of bacteriophage λ I. OR3 and autogenous negative control by repressor. *J. Mol. Biol.* **139**, 147–161 (1980).
25. Meyer, B., Maurer, R. & Ptashne, M. Gene regulation at the right operator of bacteriophage λ II. OR1, OR2, and OR3: their roles in mediating the effects of repressor and cro. *J. Mol. Biol.* **139**, 163–194 (1980).
26. Brent, R. & Ptashne, M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729–736 (1985).
27. Spencer, D.M., Wandless, T.J., Schreiber, S.L. & Crabtree, G.R. Controlling signal transduction with synthetic ligands. *Science* **262**, 1019–1024 (1993).
28. Ryan, A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).
29. Raymond, E.S. *The Cathedral and the Bazaar* (O'Reilly and Associates, San Francisco, 1999).
30. Weber, S. *The Success of Open Source* (Harvard University Press, Cambridge, Massachusetts, 2004).
31. Benford, G. The biological century. *Reason November* (1995) (Revised version: *Reason Online*, July 2001; http://reason.com/9511/BENFORDfeat.shtml.)