PERSPECTIVE

Newton, Laplace, and The Epistemology of Systems Biology

Michael L. Bittner[1] and Edward R. Dougherty[1,2]

[1]Computational Biology Division, Translational Genomics Research Institute, Phoenix, AZ, USA. [2]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA.
Corresponding author email: edward@ece.tamu.edu

Abstract: For science, theoretical or applied, to significantly advance, researchers must use the most appropriate mathematical methods. A century and a half elapsed between Newton's development of the calculus and Laplace's development of celestial mechanics. One cannot imagine the latter without the former. Today, more than three-quarters of a century has elapsed since the birth of stochastic systems theory. This article provides a perspective on the utilization of systems theory as the proper vehicle for the development of systems biology and its application to complex regulatory diseases such as cancer.

Keywords: cancer, control, epistemology, systems biology

Article

Isaac Newton published his *Principia* in 1646; Pierre-Simon Laplace published the first volume of his *Celestial Mechanics* 150 years later in 1796. Laplace's system depends on the calculus of Newton and its subsequent developments over a century and a half. Laplace did not ignore the well-developed mathematics of his day and try to develop his mechanics without it; rather, he used the relevant available tools.

Today we stand on the huge development of stochastic processes and systems theory over more than three quarters of a century. Ignoring systems theory in the development of systems biology would be analogous to Laplace trying to develop celestial mechanics with elementary algebra or Einstein trying to develop the general theory of relativity using Euclidean geometry. Whereas Laplace utilized the calculus because it was a suitable medium for the velocity, acceleration, and mass of classical mechanics, Einstein utilized Riemannian geometry because it was a suitable medium for relativistic velocity, acceleration, and mass. In both cases, formal conceptualization of the theory depended upon the availability of suitable mathematics.

Reflecting on his investigations into systems biology, in 1935 Conrad Waddington wrote, "To say that an animal is an organism means in fact two things: firstly, that it is a system made up of separate parts, and secondly, that in order to describe fully how any one part works one has to refer either to the whole system or to the other parts."[1] This was around the time that Andrey Kolmogorov was formulating a rigorous theory of continuous time random processes. Norbert Wiener was part of the rapid development of that theory and its applications in the 1930s in the United States, the Soviet Union, France, and England. In 1945, he and his physiologist collaborator, Arturo Rosenblueth, published a seminal paper in systems biology, "The mathematical formation of the problem of conduction of impulses in a network of connected excitable elements, specifically in cardiac muscle."[2] It is fitting that Wiener, the father of modern systems theory in engineering, would be the first to recognize systems theory as the natural setting for characterizing biological systems. In 1949 he declared, "Many perhaps do not realize that the present age is ready for a significant turn in the development toward far greater heights than we have ever anticipated. The point of departure may well be the recasting and unifying of the theories of control and communication in the machine and in the animal on a statistical basis."[3]

Much biological research over the past 50 years has focused on discovering sets of components required to execute the many processes necessary for cell survival and the collaboration necessary to form functioning organisms. Currently, we can identify what is likely a large percentage of the genes in complex organisms. For a portion of those, we have knowledge of some capabilities of their protein products. Our understanding of how gene products collaborate to carry out cellular processes varies considerably. In the areas of metabolism and energetics, knowledge of how the most basic building blocks of the cell are built from simple precursors and the ways in which energy is obtained to carry out cellular operations is quite detailed. Knowledge in this sphere is fairly certain due to the high degree of linearity of the operations constituting the processes. In metabolic pathways, simple substrates progress through a series of ordered, sequential chemical transformations, each mediated by a specific enzyme. These processes can be readily studied in a piecemeal fashion and the pieces assembled into a coherent whole, since each step operates on only a single or very limited set of substrates. This level of simplicity is not evident in the complex processes that constitute the wide variety of cellular activities that allow cells to develop, differentiate, and assemble into the many distinct types whose functions are required to support and maintain an organism's activities. Regulation of these activities can have many independent inputs, each capable of exerting control, and a number of parallel processes, each capable of carrying out a control process yielding the same functional result. These may be configured with feedback loops that can increase its output signal if the process experiences interference.

Many human diseases arise when there are changes in either the amount or the structure of a specific gene product in a particular cell type. Treatments of diseases arising in the metabolic domain are usually much more successful than those in other cellular function domains. A listing of inherited metabolic diseases from the Canadian Ministry of Health lists 85 different metabolic diseases for which treatments are readily available. Treatments are either essential

metabolites for those not capable of synthesizing them or foods that do not contain metabolic precursors that are toxic for those who cannot metabolize them. Whole genome sequencing of persons with diseases for which a diagnosis had not been possible has detected alterations in genes identifiable as having metabolic functions, which in some cases has allowed treatment with supplements that correct the metabolic deficiencies resulting from the altered gene. The very direct route from observation of an altered component of a metabolic process to provision of a metabolite downstream of the non-functional enzyme is a result of highly certain knowledge of the roles of the process components and a highly deterministic regulatory regime.

On the other hand, attempts to identify ways of altering cellular processes to alleviate the destruction induced by key genes influencing neurological functions,[4] or pulmonary function, cystic fibrosis transmembrane conductance regulator,[5] have been unsuccessful. Successful reversal of Huntington's disease in mouse models has recently been achieved by two groups using different strategies to reduce the amount of mutant Huntingtin protein in the neural cells, thereby avoiding intervention in the diverse processes disturbed by the protein.[6,7] With cystic fibrosis, progress has been made, not by interventions, but by using a drug cocktail capable of altering the conformation of one of the mutant forms of the cystic fibrosis gene so that it regains its normal functionality.[8] One study of Huntington's disease tested the effect of increased expression or partial loss of function of a subset of 60 genes drawn from studies that identified 234 proteins that physically interacted with mutant Huntingtin protein to determine whether changes in the behavior of these interacting genes would alter the of neurotoxicity induced by mutant Huntingtin.[9] Many modifiers involved in a wide variety of cellular processes were found that produced statistically significant, but modest toxicity modification. This is a striking example of the hazard that uncertainty poses to a researcher attempting to develop an intervention strategy. The more broadly a gene capable of inducing a pathological cellular state perturbs the normal cellular processes, the harder it is to discover a successful strategy that acts on any gene save the pathological gene itself.

Efforts to control pathological cell processes by interventions at fixed points in specific cellular processes have been enthusiastically applied in attempts to develop oncology drugs over the last two decades, the result being a very limited number of drugs producing high curative rates. The majority of drugs produce temporary reductions in tumor abundance over periods of one to six months, followed by active tumor proliferation. This was first thought to result from misalignment of the drug with the tumor's molecular characteristics, but it was later observed that even a frequently used drug such as gefitinb does not provide a significant increase in overall survival, even for patients having the most favorable molecular characteristics for response.[10] The failure of the drug vemurafinib, which targets a mutated form of the BRAF kinase, to produce a durable response in melanoma in spite of a very strong initial response of melanoma tumors to the drug has launched a widespread investigation into how these tumors overcome the drug's interventive effects.[11–15] As witnessed for other targets, such as members of the EGFR family of receptor tyrosine kinases,[16] a wide variety of resistance mechanisms can be deployed by cancer cells. These include induction of alternate processes that provide overlapping activation of many of the same processes that the target normally activates, outgrowth of either newly acquired mutations or a minor population of tumor cells already bearing a mutated component that activates the processes activated by the target, the presence of feedback loops that provide target activity levels higher than the targeting drug can suppress, induction of a resistance producing alternate process by RTK ligands produced by stromal cells in the tumor, and other mechanisms. Reviewing these complexities, Yosef Yarden, a distinguished researcher with considerable experience in the biology of the EGFR family, argues that progress in developing effective treatment strategies requires that the knowledge of the various networks involved in and activated by this gene family needs to be assembled into a system view of the overall process, to allow knowledge of the various mutations and gene product levels in a particular patient to be used to specify what single or multiple drug perturbations will provide an effective therapy.[16]

In sum, biological modeling, and the translational medicine consequent to it, must handle the parallelism and redundancy required for system efficiency and survivability. This problem has been faced before by

engineers and scientists in complex system analysis, albeit, not as complex as biological systems. Only via interwoven regulation can a system be sufficiently fault tolerant to survive in a rapidly changing interactive environment. We confront the modeling and control of systems capable of autonomous reconfiguration. This problem has been faced by engineers and scientists in complex system analysis since the 1930s.

Reflecting on the development of control systems to regulate autonomously reconfigurable systems, in the 1960s, stochastic control theorist Vladimir Pugachev wrote, "The simplest systems of this type, which incorporate elements for automatically adjusting particular parameters according to an analysis of input and output data, are called *self-adjusting* systems. Complex systems of this kind are capable of adapting themselves completely at each instant to the results of their analysis of external conditions and previous performance. These are said to be *self-organizing*. It is quite clear that no theory of error under average operating conditions is adequate for the design of self-adjusting and self-organizing systems. A special theory is required which will solve the complex problems involved in processing the input data and utilizing it to best advantage in any particular case. Both problems can be tackled by the modern theory of optimal systems."[17] If we define biology as the "study of organisms, physical systems capable of retaining and utilizing information to execute processes that utilize available energy to organize matter for facilitation of their own persistence and reproduction",[18] then biology clearly falls within the purview of Pugachev's characterization.

Today, we possess much greater knowledge of systems theory and orders of magnitude greater computational power than were available to Pugachev's generation, but to take advantage of this greater knowledge and power it is first necessary to overcome a challenging problem: how to embed biological knowledge in stochastic systems theory. It would be foolhardy to underestimate the difficulty of the problem. Indeed, consider the decades it took to develop applications in electrical, mechanical, and computer engineering. Biological systems present a higher hurdle. Nonetheless, when one considers the control necessary to land a man on the moon, less than 40 years after Kolmogorov's fundamental paper, one should not face the development of biological systems theory with trepidation, especially since we have much more knowledge of nonlinear systems than was possessed in the 1960s and our technical apparatus dwarfs what was available then. But to achieve the goal, one must begin the trek. It begins with biology becoming a science of stochastic systems.

Work over the last decade on the external control of gene regulatory networks modeled as Markov chains provides a peek into a future biomedicine embedded in stochastic systems theory. Model-based design of intervention strategies using stochastic control was first applied to gene regulatory networks over a finite horizon (finite time window).[19] Following some further early work with finite-horizon control, attention shifted to optimal infinite-horizon control, the intent being to drive the long-run behavior of the network towards desirable phenotypes.[20] Since then attention has shifted to dealing with practical issues.

The size and complexity of gene regulatory networks creates what may be the two most challenging problems for stochastic control. The first is computational complexity. Policy design algorithms, whether they involve dynamic programming or matrix manipulations, are quickly overwhelmed by state spaces arising from even a small gene set. One way to address computational limitations is to use greedy control policies that forego full optimization;[21,22] however, these are quickly overcome by the exponential growth in the state space relative to the number of genes. Another approach is network reduction, where a compression algorithm is used to delete genes or network states that, based on some measure, are not important relative to the control objective.[23–25]

A second challenging issue is model uncertainty. Owing to biological complexity and experimental limitations, model uncertainty is virtually unavoidable. Control policies need to be robust with respect to uncertain modeling assumptions.[26–29] Specifically, rather than design a control policy that is optimal for a particular network, design one that is optimal (in some sense) across an uncertainty class of networks. Moreover, rather than depend on data alone, apply prior biological knowledge to constrain the uncertainty class. This presumes a method of taking biological knowledge and data, say, in the form of pathways and steady-state expression data, and producing networks

consistent with both the knowledge and the data.[30,31] As with any scientific experiment, it is important to follow an experimental protocol suitable to the goal of designing of a stochastic control policy, not merely perusing haphazard data.[18]

Other practical issues abound, such as limiting drug dosage,[32] allowing time for recovery following treatment,[33,34] and constraining optimality so as not to induce phenotypes that, while not known to be pathological from the standpoint of the disease of interest, may nonetheless be undesirable.[35] As medicine joins other modern engineering disciplines and *ad hoc* operational regimes are replaced by optimized procedures based on stochastic systems theory,[36] a host of other practical issues will have to be modeled and mathematically addressed. Beyond the theory, dynamic experimental protocols must be developed for both model design and validation, and the concomitant statistical issues addressed.

Altogether this will be a monumental effort requiring a transformation of biomedical thinking almost as radical as that from medieval science to Newtonian mechanics. While it would certainly be naive to think that this transformation will be easy, it would be equally naïve to think that one could build a science of biological systems and translate that science into medical treatment of extraordinarily complex regulatory diseases such as cancer using sophomore mathematics augmented with powerful search engines.

Author Contributions

Wrote the first draft of the manuscript: MLB, ERD. Contributed to the writing of the manuscript: MLB, ERD. Agree with manuscript results and conclusions: MLB, ERD. Jointly developed the structure and arguments for the paper: MLB, ERD. Made critical revisions and approved final version: MLB, ERD. All authors reviewed and approved of the final manuscript.

Funding

Author(s) disclose no funding sources.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Waddington CH. *How Animals Develop*. London: Allen and Unwin. 1935.
2. Wiener N, Rosenblueth A. The mathematical formation of the problem of conduction of impulses in a network of connected excitable elements, specifically in cardiac muscle. *Archives Institute Cardiology, Mexico*. 1946;16:204–65.
3. Wiener N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge: MIT Press. 1949.
4. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell*. 1993;72(6):971–83.
5. Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*. 1989;245(4922):1066–73.
6. Kordasiewicz HB, Stanek LM, Wancewicz EV, et al. Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron*. 2012;74(6):1031–44.
7. Di Pardo A, Maglione V, Alpaugh M, et al. Ganglioside GM1 induces phosphorylation of mutant huntingtin and restores normal motor behavior in Huntington disease mice. *Proc Natl Acad Sci U S A*. 2012;109(9): 3528–33.
8. Yu W, Kim Chiaw P, Bear CE. Probing conformational rescue induced by a chemical corrector of F508del-cystic fibrosis transmembrane conductance regulator (CFTR) mutant. *J Biol Chem*. 2011;286(28):24714–25.
9. Kaltenbach LS, Romero E, Becklin RR, et al. Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genetics*. 2007;3(5):e82.
10. Azuma K, Okamoto I, Kawahara A, et al. Association of the expression of mutant epidermal growth factor receptor protein as determined with mutation-specific antibodies in non-small cell lung cancer with progression-free survival after gefitinib treatment. *J Thorac Oncol*. 2012;7(1):122–7.
11. Poulikakos PI, Rosen N. Mutant BRAF melanomas—dependence and resistance. *Cancer Cell*. 2011;19(1):11–5.
12. Basile KJ, Abel EV, Aplin AE. Adaptive upregulation of FOXD3 and resistance to PLX4032/4720-induced cell death in mutant B-RAF melanoma cells. *Oncogene*. 2012;31(19):2471–9.
13. Heidorn SJ, Milagre C, Whittaker S, et al. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell*. 2010;140(2):209–21.
14. Corcoran RB, Dias-Santagata D, Bergethon K, et al. BRAF gene amplification can promote acquired resistance to MEK inhibitors in cancer cells harboring the BRAF V600E mutation. *Sci Signal*. 2010;3(149):ra84.
15. Straussman R, Morikawa T, Shee K, et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature*. 2012;487(7408):500–4.
16. Yarden Y, Pines G. The ERBB network: at last, cancer therapy meets systems biology. *Nat Rev Cancer*. 2012;12(8):553–63.

17. Pugachev VS. *Theory of Random Functions and Its Application to Control Problems*. Oxford: Pergamon Press. 1965.

18. Dougherty ER, Bittner ML. *Epistemology of the Cell: A Systems Perspective on Biological Knowledge*. IEEE Press Series on Biomedical Engineering. New York: John Wiley. 2011.

19. Datta A, Choudhary A, Bittner ML, Dougherty ER. External control in Markovian genetic regulatory networks. *Machine Learning*. 2003; 52(1–2):169–91.

20. Pal R, Datta A, Dougherty ER. Optimal infinite horizon control for probabilistic Boolean networks. *IEEE Trans Signal Proc*. 2006;54(6 Pt 2): 2375–87.

21. Vahedi G, Faryabi B, Chamberland J-F, Datta A, Dougherty ER. Intervention in gene regulatory networks via a stationary mean-first-passage-time control policy. *IEEE Tran Biomed Eng*. 2008;55(10):2319–31.

22. Qian X, Ivanov I, Ghaffari N, Dougherty ER. Intervention in gene regulatory networks via greedy control policies based on long-run behavior. *BMC Systems Bio*. 2009;3:61.

23. Qian X, Ghaffari N, Ivanov, I, Dougherty ER. State reduction for network intervention with probabilistic Boolean networks. *Bioinformatics*. 2010;26(24):3098–194.

24. Ivanov I, Simeonov P, Ghaffari N, Qian X, Dougherty ER. Selection policy induced reduction mappings for Boolean networks. *IEEE Trans on Sig Proc*. 2010;58(9):4871–82.

25. Ghaffari N, Ivanov I, Qian X, Dougherty ER. A CoD-based reduction algorithm for designing stationary control policies on Boolean networks. *Bioinformatics*. 2010;26:1556–63.

26. Pal R, Datta A, Dougherty ER. Robust intervention in probabilistic Boolean networks. *IEEE Trans Sig Proc*. 2008;56(3):1280–94.

27. Pal R, Datta A, Dougherty ER. Bayesian robustness in the control of gene regulatory networks. *IEEE Trans Sig Proc*. 2009;57(9):3667–78.

28. Pal R, Bhattacharya S. Characterizing the effect of coarse-scale PBN modeling on dynamics and intervention performance of genetic regulatory networks represented by stochastic master equation. *IEEE Trans Sig Proc*. 2010;58(6):3341–3351.

29. Pal R, Bhattacharya S, Caglar MU. Robust approaches for genetic regulatory network modeling and intervention: a review of recent advances. *IEEE Sig Proc Mag*. 2012;29(1):66–76.

30. Layek R, Datta A, Dougherty ER. From biological pathways to regulatory networks. *Molecular Bio Syst*. 2011;7(3):843–51.

31. Knight J, Datta A, Dougherty ER. Generating stochastic gene regulatory networks consistent with pathway information and steady-state behavior. *IEEE Trans Biomed Eng*. 2012;59(6):1701–10.

32. Faryabi B, Vahedi G, Chamberland J-F, Datta A, Dougherty ER. Optimal constrained stationary intervention in genetic regulatory networks. *EURASIP Journal on Bioinformatics and Systems Biology*. 2008. doi:10.1155/2008/620767.

33. Vahedi G, Faryabi B, Chamberland J-F, Datta A, Dougherty ER. Optimal intervention strategies for cyclic therapeutic methods with fixed-length duration of effect. *IEEE Trans Biomed Eng*. 2009;56(2):281–91.

34. Yousefi MR, Datta A, Dougherty ER. Optimal control strategies for therapeutic methods with fixed-length duration of drug effectiveness. *IEEE Trans Sig Proc*. 2012;60(9):4930–44.

35. Qian X, Dougherty ER. Intervention in gene regulatory networks via phenotypically constrained control policies based on long-run behavior. *IEEE/ACM Trans Comp Bio and Bioinformatics*. 2012;9(1):123–36.

36. Dougherty ER. Translational science: epistemology and the investigative process. *Current Genomics*. 2009;10(2):102–9.