The Promise and Challenge of Digital Biology

Mark E Minie[1] and Ram Samudrala[2]*

[1]Bioengineering Department, University of Washington, USA
[2]Microbiology Department, University of Washington, USA

Introduction

The era of Digital Biology began in 2010 with the "rebooting" of a bacterial cell using a synthetic DNA genome created from a digital template stored on a computer [1]. With this event, the creation of *Mycoplasma laboratorium* (nicknamed "Synthea"), came the first complete proof that DNA was the true software of life. Cells could be simulated digitally and the simulations could be tested against reality by reprograming cytoplasm with synthetic genomes generated from the digital DNA sequences driving those simulations. This in turn has created the expectation and promise that a deeper understanding of cellular function and thus life itself could be achieved on an infinite iterative loop of computer modeling and chemical synthesis (Figure 1) [2].

Key components of the digital biology loop are 1) a detailed digital mapping of living systems and their biomolecular parts and the interactions of such parts-biodigitization, 2) accessible databases containing/managing this biodata, 3) computer simulation algorithms of cells driven by digital DNA sequences encoding the biomolecular parts and interactions-biosimulation, 4) laboratory technologies to deeply analyze the resulting synthetic cells-biolab-and finally and centrally 5) the digital biological converter (digital bioconverter for short). In these early days of digital biology, each of these components presents exciting bioengineering, bioscience and biomedical challenges.

Biodigitization and Biodata

Every aspect of Earth's biosphere is currently being digitized, from the molecular to the planetary levels and the data entered into an ever-growing collection of biologically oriented Internet databases-a mirror backup image of terrestrial life is literally being created (Figure2A)[3]. The digital acquisition of DNA sequences of phage, viruses, bacteria and human cells, the 3D structures of biomolecules as well as the detailed cellular structures and tissue and organ architecture has been underway in earnest for more than two decades now. With the advent of molecular imaging, electronic medical records, and "Big Data" [4,5], every aspect of individual organisms, populations, and ecosystems are now also being fed into the Internet based DataStream. Online biological databases are also on track for doubling every 5 years [3,6]
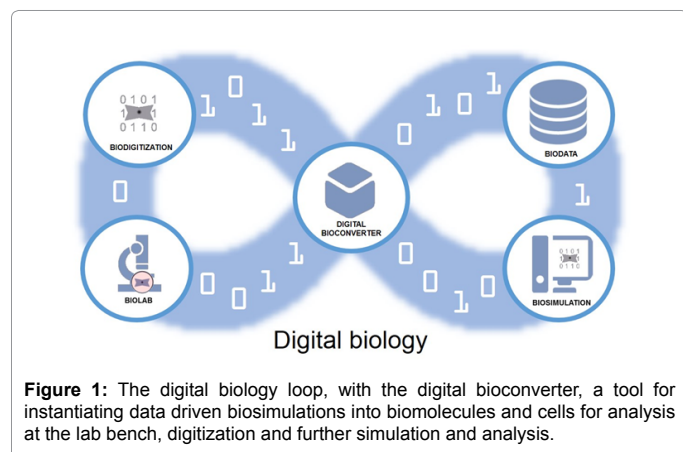

Figure 1: The digital biology loop, with the digital bioconverter, a tool for instantiating data driven biosimulations into biomolecules and cells for analysis at the lab bench, digitization and further simulation and analysis.

(Figure 2B). The sheer volume of such data now threatens to overtake current data storage and search technology, and may require the development of novel technologies, including nucleic acid based data storage [7] and quantum computing [8].

Biosimulation and Biolab

While a number of simulations of cells have been developed over the last two decades [9–11], so far only one specifically driven by a digitized DNA genome from a real cell [12], Mycoplasma genitalium, has been published. The advent of synthetic biology tools such as Tinkercell [12–14] and database driven animation [15] provide strong starts for the tools that will be needed. Detailed electronic images of biomolecules useful in simulations are already readily available from such sources as PubChem and Biosystems [6]. On the other side of the issue, biolab tools for analyzing and manipulating synthetically produced living cells are arising with breathtaking speed. Molecular imaging of structures within living cells is now possible [16–19], and even the direct physical manipulation of cellular components using optical tweezers is a routine technique [20]. Additive manufacturing tools now make the construction of artificial biofilms and organs for research a reality [21–28] (Figure 3).

Digital Bioconverter

The key "gadget" in digital biology, the digital bioconverter (Figure 4), currently exists as a prototype [2] and will likely eventually evolve
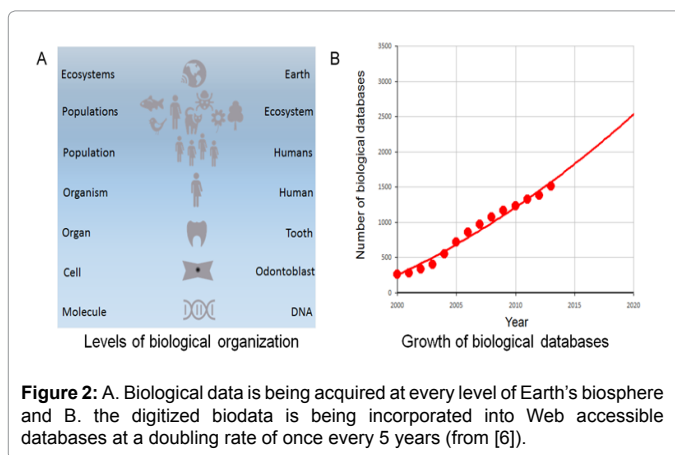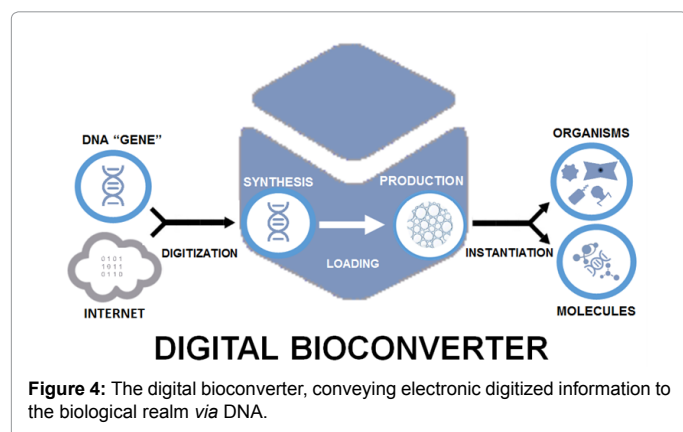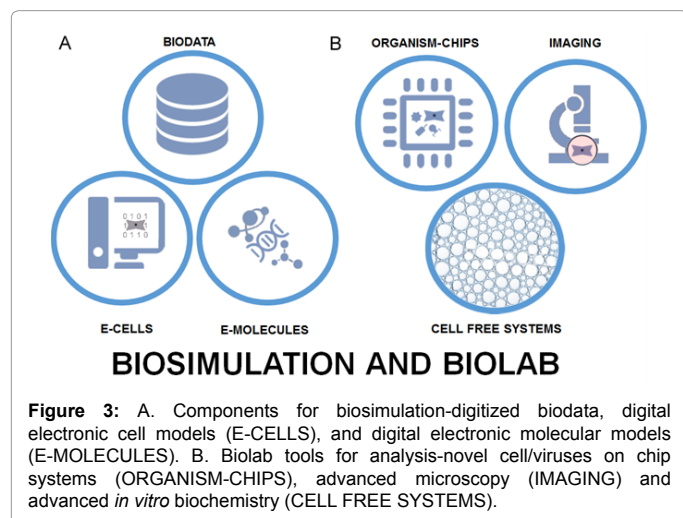

Figure 2: A. Biological data is being acquired at every level of Earth's biosphere and B. the digitized biodata is being incorporated into Web accessible databases at a doubling rate of once every 5 years (from [6]).

Corresponding author: Ram Samudrala, Microbiology Department, University of Washington, USA, Tel: 1-206-732-6122; E-mail: ram@compbio.washington.edu

Figure 3: A. Components for biosimulation-digitized biodata, digital electronic cell models (E-CELLS), and digital electronic molecular models (E-MOLECULES). B. Biolab tools for analysis-novel cell/viruses on chip systems (ORGANISM-CHIPS), advanced microscopy (IMAGING) and advanced *in vitro* biochemistry (CELL FREE SYSTEMS).

Figure 4: The digital bioconverter, conveying electronic digitized information to the biological realm *via* DNA.

into a miniaturized commercially produced laboratory instrument. Such a system would allow the convenient production of cells, viruses and biological molecules directly from digitized gene encoding DNA sequences, and eventually could be as central to basic bioscience research as automated DNA sequencers are today. Significant challenges must be met before this is realized, however. While synthesis of large genomes is now possible, it remains complex and expensive. An alternative to cellular transformation could be realized *via* cell free systems-DNA could be loaded into such systems and then drive the production of biomolecules or organisms through a further instantiation step-possibly through reconstitution of cells and viruses (organisms) from *in vitro* systems [29–32] or the cell free synthesis of biomolecules.

Grand Synthesis

Increasingly available digitized biodata coupled with advanced biosynthetic synthesis are leading to a new era of biology where electronic digital simulations can be converted to cells and biological molecules-the era of digital biology. While several major engineering and computing challenges must be tackled, these are not insurmountable and are the objects of vigorous technological innovation. In particular, the development of a standardized and commercially available bioconversion device will be critical, and with such a device in eventual widespread use a rapid cycle of model driven understanding of biological systems will get underway. Such a device and the concept of digital biology will have applications in many fields, including astrobiology [33–35], medicine [36–38], nanotechnology [39], bioinformatics

[36,37,40–47], drug repurposing [48] and pharmacoengineering [49–51], while presenting the promise of placing bioengineering and biomedicine on a Moore's Law-like curve of exponentially increasing understanding and providing exquisite control of living systems.

Reference

1. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329: 52-56.

2. Venter JC(2013) Life at the Speed of Light: From the Double Helix to the Dawn of Digital Life. Viking Adult

3. Fernández-Suárez XM, Galperin MY (2013) The 2013 Nucleic Acids Research Database Issue and the online molecular biology database collection. Nucleic Acids Res 41: D1-7.

4. Daves, K (2011) Big Data, BGI and GigaScience. Bio-IT World 1-4.

5. Ratner B (2011) Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeondling and Analysis of Big Data. CRC Press.

6. Abigail Acland R, Agarwala, Tanya Barrett, Jeff Beck DAB, Colleen Bollin, et al. (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 41: D8–D20

7. Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, et al. (2013) Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature 494: 77-80.

8. Harris SA, Kendon VM (2010) Quantum-assisted biomolecular modelling. Philos Trans A Math Phys Eng Sci 368: 3581-3592.

9. Ishii N, Robert M, Nakayama Y, Kanai A, Tomita M (2004) Toward large-scale modeling of the microbial cell for computer simulation. J Biotechnol 113: 281-294.

10. Klein MT, Hou G, Quann RJ, Wei W, Liao KH, et al. (2002) BioMOL: a computer-assisted biological modeling tool for complex chemical mixtures and biological processes at the molecular level. Environ Health Perspect 110 Suppl 6: 1025-1029.

11. Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, et al. (1999) E-CELL: software environment for whole-cell simulation. Bioinformatics 15: 72-84.

12. Karr JR, Sanghvi JC, Macklin DN, Gutschow M V, et al. (2012) A whole-cell computational model predicts phenotype from genotype. Elsevier 150: 389–401.

13. Chandran D, Bergmann FT, Sauro HM (2010) Computer-aided design of biological circuits using TinkerCell. Bioeng Bugs 1: 274-281.

14. Chandran D, Sauro HM (2012) Hierarchical modeling for synthetic biology. ACS Synth Biol 1: 353-364.

15. McGill G (2008) Molecular movies... coming to a lecture near you. Cell 133: 1127-1132.

16. Grammel M, Hang HC (2013) Chemical reporters for biological discovery. Nat Chem Biol 9: 475-484.

17. Chung S, Takizawa PA (2011) In vivo visualization of RNA using the U1A-based tagged RNA system. Methods Mol Biol 714: 221-235.

18. Courty S, Dahan M (2013) Tracking individual intracellular proteins using quantum dots. Cold Spring Harb Protoc 2013.

19. O'Rourke NA, Meyer T, Chandy G (2005) Protein localization studies in the age of 'Omics'. Curr Opin Chem Biol 9: 82-87.

20. Oddershede LB (2012) Force probing of individual molecules inside the living cell is now a reality. Nat Chem Biol 8: 879-886.

21. Lee JB, Sung JH (2013) Organ-on-a-chip technology and microfluidic whole-body models for pharmacokinetic drug toxicity screening. Biotechnol J .

22. Sung JH, Yu J, Luo D, Shuler ML, March JC (2011) Microscale 3-D hydrogel scaffold for biomimetic gastrointestinal (GI) tract model. Lab Chip 11: 389-392.

23. Neuži P, Giselbrecht S, Länge K, Huang TJ, Manz A (2012) Revisiting lab-on-a-chip technology for drug discovery. Nat Rev Drug Discov 11: 620-632.

24. Ghaemmaghami AM, Hancock MJ, Harrington H, Kaji H, Khademhosseini A (2012) Biomimetic tissues on a chip for drug discovery. Drug Discov Today 17: 173-181.

25. Sung JH, Esch MB, Prot JM, Long CJ, Smith A, et al. (2013) Microfabricated

mammalian organ systems and their integration into models of whole animals and humans. Lab Chip 13: 1201-1212.

26. Atala A (2012) Regenerative medicine strategies. J Pediatr Surg 47: 17–28.

27. Mironov V, Kasyanov V, Markwald RR (2011) Organ printing: from bioprinter to organ biofabrication line. Curr Opin Biotechnol 22: 667-673.

28. Mertz L (2013) New world of 3-d printing offers "completely new ways of thinking": q&a with author, engineer, and 3-d printing expert hod lipson. IEEE Pulse 4: 12-14.

29. Shin J, Jardine P, Noireaux V (2012) Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. ACS Synth Biol 1: 408-413.

30. Abdoli A, Soleimanjahi H, Kheiri MT, Jamali A, Sohani H, et al. (2013) Reconstruction of H3N2 influenza virus based virosome in-vitro. Iran J Microbiol 5: 166-171.

31. Ge X, Luo D, Xu J (2011) Cell-free protein expression under macromolecular crowding conditions. PLoS One 6: e28707.

32. Sokolova E, Spruijt E, Hansen MM, Dubuc E, Groen J, et al. (2013) Enhanced transcription rates in membrane-free protocells formed by coacervation of cell lysate. Proc Natl Acad Sci U S A 110: 11692-11697.

33. Blumberg BS (2011) Astrobiology, space and the future age of discovery. Philos Trans A Math Phys Eng Sci 369: 508-515.

34. Goldman AD, Samudrala R, Baross JA (2010) The evolution and functional repertoire of translation proteins following the origin of life. Biol Direct 5: 15.

35. Goldman AD, Baross JA, Samudrala R (2012) The enzymatic and metabolic capabilities of early life. PLoS One 7: e39912.

36. Horst JA, Laurenzi A, Bernard B (2012) Computational Multitarget Drug Discovery. Polypharmacology Drug Discov 263–301.

37. Jenwitheesuk E, Horst JA, Rivas KL, Van Voorhis WC, Samudrala R (2008) Novel paradigms for drug discovery: computational multitarget screening. Trends Pharmacol Sci 29: 62-71.

38. Jenwitheesuk E, Samudrala R (2005) Identification of potential multitarget antimalarial drugs. JAMA 294: 1490-1491.

39. Doll TA, Raman S, Dey R, Burkhard P (2013) Nanoscale assemblies and their biomedical applications. J R Soc Interface 10: 20120740.

40. Sun X, Vilar S, Tatonetti NP (2013) High-throughput methods for combinatorial drug discovery. Sci Transl Med 5: 205rv1.

41. Yu J, Wang J, Lin W, Li S, Li H, et al. (2005) The Genomes of Oryza sativa: a history of duplications. PLoS Biol 3: e38.

42. Hung LH, Samudrala R (2003) PROTINFO: Secondary and tertiary protein structure prediction. Nucleic Acids Res 31: 3296-3299.

43. Samudrala R (2002) Modeling genome structure and function. Pure Appl Chem 74: 907–914.

44. Hung LH, Ngan SC, Liu T, Samudrala R (2005) PROTINFO: new algorithms for enhanced protein structure predictions. Nucleic Acids Res 33: W77-80.

45. Yuriev E, Ramsland PA (2013) Latest developments in molecular docking: 2010-2011 in review. J Mol Recognit 26: 215-239.

46. Horst J, Samudrala R (2009) Diversity of protein structures and difficulties in fold recognition: the curious case of protein G. F1000 Biol Rep 1: 69.

47. Ou-Yang SS, Lu JY, Kong XQ, Liang ZJ, Luo C, et al. (2012) Computational drug discovery. Acta Pharmacol Sin 33: 1131-1140.

48. Xie L, Xie L, Kinnings SL, Bourne PE (2012) Novel computational approaches to polypharmacology as a means to define responses to individual drugs. Annu Rev Pharmacol Toxicol 52: 361-379.

49. Papworth M, Kolasinska P, Minczuk M (2006) Designer zinc-finger proteins and their applications. Gene 366: 27-38.

50. Chen Y, Varani G (2013) Engineering RNA-binding proteins for biology. FEBS J 280: 3734-3754.

51. Jinek M, East A, Cheng A, Lin S, Ma E, et al. (2013) RNA-programmed genome editing in human cells. Elife 2: e00471.