Title

Perspectives for self-driving labs in synthetic biology

Authors

Martin, Hector G
Radivojevic, Tijana
Zucker, Jeremy
et al.

Publication Date

2023-02-01

DOI

10.1016/j.copbio.2022.102881

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Perspectives for self-driving labs in synthetic biology

Hector G Martin[1,2,3,4], Tijana Radivojevic[1,2,3], Jeremy Zucker[16], Kristofer Bouchard[1,5,6], Jess Sustarich[3,15], Sean Peisert[5,8], Dan Arnold[7], Nathan Hillson[1,2,3], Gyorgy Babnigg[2,13], Jose M Marti[1,2,3,9], Christopher J Mungall[1], Gregg T Beckham[2,17], Lucas Waldburger[18], James Carothers[12], ShivShankar Sundaram[10,11], Deb Agarwal[5], Blake A Simmons[1,2,3], Tyler Backman[1,3], Deepanwita Banerjee[1,3], Deepti Tanjore[1,2,14], Lavanya Ramakrishnan[5] and Anup Singh[3,10]

Self-driving labs (SDLs) combine fully automated experiments with artificial intelligence (AI) that decides the next set of experiments. Taken to their ultimate expression, SDLs could usher a new paradigm of scientific research, where the world is probed, interpreted, and explained by machines for human benefit. While there are functioning SDLs in the fields of chemistry and materials science, we contend that synthetic biology provides a unique opportunity since the genome provides a single target for affecting the incredibly wide repertoire of biological cell behavior. However, the level of investment required for the creation of biological SDLs is only warranted if directed toward solving difficult and enabling biological questions. Here, we discuss challenges and opportunities in creating SDLs for synthetic biology.

Addresses
[1] Lawrence Berkeley National Laboratory, Biological Systems and Engineering Division, Berkeley, CA, United States
[2] Department of Energy, Agile BioFoundry, Emeryville, CA, United States
[3] Joint BioEnergy Institute, Emeryville, CA, United States
[4] BCAM, Basque Center for Applied Mathematics, Bilbao, Spain
[5] Lawrence Berkeley National Laboratory, Scientific Data Division, Berkeley, CA, United States
[6] Helen Wills Neuroscience Institute and Redwood Center for Theoretical Neuroscience, Berkeley, CA, United States
[7] Lawrence Berkeley National Laboratory, Energy Storage and Distributed Resources Division, Berkeley, CA, United States
[8] University of California, Davis, Department of Computer Science, Davis, CA, United States
[9] Global Security Computing Applications Division, Lawrence Livermore National Laboratory, Livermore, CA, United States
[10] Engineering Directorate, Lawrence Livermore National Laboratory, Livermore, CA, United States
[11] Center for Bioengineering, Lawrence Livermore National Laboratory, Livermore, CA, United States
[12] Department of Chemical Engineering, Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, WA, United States
[13] Biosciences Division, Argonne National Laboratory, Argonne, IL, United States
[14] Advanced Biofuels and Bioproducts Process Development Unit, Lawrence Berkeley National Laboratory, Berkeley, CA, United States
[15] Biomaterials and Biomanufacturing Division, Sandia National Laboratories, Livermore, CA, United States
[16] Earth and Biological Sciences Division, Pacific Northwest National Laboratories, Richland, WA, United States
[17] Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO 80401, United States
[18] Department of Bioengineering, University of California, Berkeley, CA, United States

Corresponding author: Martin, Hector G (hgmartin@lbl.gov)

Current Opinion in Biotechnology 2023, **79**:102881

This review comes from a themed issue on **Systems Biology**

Edited by **Howard Salis**

For complete overview of the section, please refer to the article collection, "Systems Biology (2023)"

Available online 3 January 2023

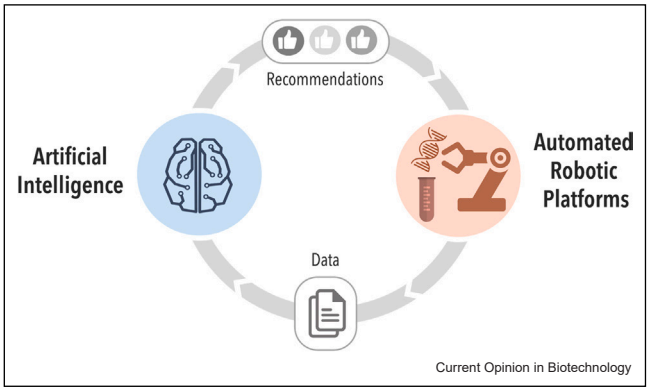
https://doi.org/10.1016/j.copbio.2022.102881

What is a self-driving lab?

Self-driving labs (SDLs), or autonomous experimentation, combine robotics for automated experiments and data collection, with artificial intelligence (AI) systems that use these data to recommend follow-up experiments [1–3] (Fig. 1). These recommendations potentially involve not just the conditions and parts to be used for the next experiment, but also which underlying hypothesis to test.

A possible example of a SDL in synthetic biology could involve a DNA assembly microfluidic chip that automatically produces variants of a given pathway producing a metabolite of interest (e.g. the biofuel precursor

Figure 1



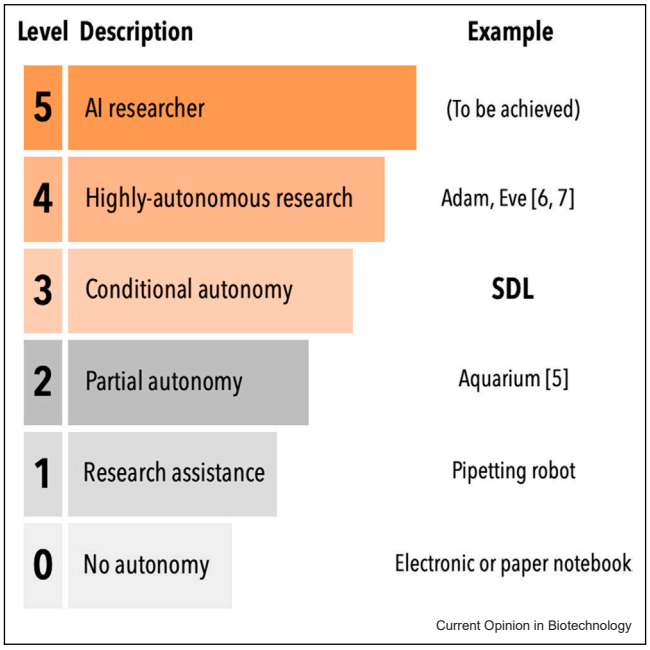
SDLs combine automated robotic platforms and data collection with AI that processes these data to decide the next set of experiments to perform and, potentially, which hypotheses and theories to test.

bisabolene), transforms them into a host (e.g. a bacteria such as *E. coli*, *P. putida*, or *R. toruloides*), and is able to culture this host and measure the corresponding bisabolene production. This automated experiment setup would be coupled with an AI recommendation engine that takes these experimental data and proposes different pathway variants with the goal of maximizing bisabolene production. A conceivable expansion could add the ability to replace specific genetic parts beyond the bisabolene pathway, so as to have an effect on precursor supply. The hypotheses generated by the AI in terms of recommended pathway variants and gene edits would be tested by the automated microfluidic chip in the next cycle.

The SDL concept requires full autonomy from humans. A partially automated system, or one that requires human intervention to finish the cycle of experimentation/planning is not, rigorously speaking, a SDL. Full automation for the cycle is not a whimsical requirement, but rather enables the full potential of SDLs. Completely automated systems can reach duty cycles (e.g. 24/7/365 operation), experiment-to-experiment reproducibility, and efficiency that are unattainable by humans. Furthermore, they are potentially linearly scalable (e.g. simply acquire more copies of the equipment), and, as a consequence, can produce large amounts of high-quality data and metadata. Such large volumes of high-quality data can make AI systems particularly effective and insightful: artificial neural networks, for example, are known to be most effective once a certain threshold of training data is available. These benefits will be critical to produce improved scientific understanding, and significantly decrease time to the desired bioproducts.

However crucial, the requirement of full autonomy may be too stringent for the current state of technology, so it is useful to consider intermediate steps toward the realization of SDLs. Similar considerations have moved the

Figure 2



SDLs are level-3 autonomy systems. Autonomy levels for SDLs describe the degree of independence from human intervention. At level 0, all experimental design and execution, as well as data capture, is handled by humans. At level 1, some repetitive tasks are outsourced to robots. Level 2 requires systematic digital description of protocols and experiments, as well as machine-interpretable data, such as in the laboratory work planner Aquarium [5]. Level 3 involves the closed Design–Build–Test–Learn cycles that can be considered the minimum requisite for a SDL, along with interpretations of routine analyses and flagging anomalies for humans to handle. Level 4 involves robotic protocol execution and routine data analyses, as in 'Adam' and 'Eve' [6,7], with humans involved only as setting goals and plans (i.e. SDL works as a lab assistant to humans). At level 5, humans just set goals and receive results (i.e. SDL behaves as investigator and human as manager).
Adapted from Beal and Rogers [4].

car industry to entertain the concept of 'Degrees of autonomy' in self-driving cars. For this reason, a similar set of 'Autonomy levels' has been proposed to both describe the current technological capabilities and incentivize the gradual development into fully autonomous systems [4] (see Fig. 2). In practical terms, systems displaying an autonomy level of three or above can be considered SDLs, since they display closed Design–Build–Test–Learn loops.

Some examples of existing self-driving labs

SDLs are not unattainable fantasies: there are indeed several published examples, although focused on narrow tasks. In chemistry and materials science, SDLs are enjoying an upsurge in popularity, and several examples are now available. In biology, there are some budding examples, which show the promise of this approach.

In chemistry [3] and material sciences [2], the maturity of automation platforms and the availability of machine learning (ML) methods has enabled the creation of several SDLs or almost fully automated processes. For example, Granda et al. [8] developed a platform that explores the chemical space using an organic synthesis robot combined with an ML model to predict reactivity of possible reagent combinations. More recently, Christensen et al. [9] developed an automated closed-loop system for parallel process optimization in reactors to optimize the yield of a stereoselective Suzuki–Miyaura cross-coupling reaction. Wang et al. [10] developed a self-optimizing millifluidic reactor for scaling the manufacturing of nanomaterials with improved optical properties. In material sciences, Macleod et al. used the modular robotic platform *Ada* capable of autonomously optimizing the hole mobility of the materials commonly used in perovskite solar cells and consumer electronics [11], as well as discovering new synthesis conditions for optimized conductivities and processing temperatures for palladium films [12]. Robotics coupled with Bayesian optimization were used in multiple cases: autonomous synthesis and resistance minimization of thin films [13], optimizing mechanical properties of structures for a given application [14], improving adhesive formulations [15], achieving targeted 3D print features in additive manufacturing [16], discovering novel battery electrolytes [17], and search for photocatalyst mixtures with improved activity for hydrogen production from water [18].

Biology saw the first published closed-loop systems for scientific discovery in the form of *Adam*, a robot scientist that determined gene function through gene deletion and auxotrophic experiments in *S. cerevisiae* [6]. *Eve* followed for the repurposing of drugs, identifying an angiogenesis-inhibiting anticancer drug for antimalarial use [7]. More recently, Si et al. [19] developed an automated platform for multiplex genome-scale engineering in *S. cerevisiae*, Hamedirad et al. [20] used the BioAutomata fully automated platform to optimize promoter choice in lycopene-producing *E. coli*, and Kanda et al. [21] used an autonomous robotic system to find the optimal conditions for inducing stem cell differentiation into retinal pigment epithelial cells.

While funding for SDLs is still limited, there are instances from the US National Science Foundation (NSF), the Canadian National Research Council, and (Defense Advanced Research Projects Agency) DARPA.

The special case of biology

SDLs present unique opportunities and challenges in biology, as compared with other disciplines in which they have been deployed.

A unique opportunity is the collection of the cellular instructions in a single repository (genomic DNA) that can now be easily manipulated via gene-editing techniques or evolutionary approaches. In material sciences, the Young's modulus or hole mobility of a material depends on a variety of structural and chemical elements that are distributed over the material, and can be complicated to locate and modify. In biology, a cell's phenotype is determined primarily by a combination of its environment and its genome. The genome's capability to encode an incredibly varied set of phenotypes is showcased by the fantastic diversity provided by evolution on Earth over the last three billion years. These phenotypes range from metabolic adaptation to extreme environments, carbon capture from atmosphere, production of valuable chemicals and bioproducts, and multicellular coordination, to the emergence of consciousness and intelligence. Furthermore, the genome is now more accessible than ever before through recent advances in CRISPR-enabled gene-editing tools [22], and evolutionary approaches comprising, for example, targeted mutagenesis and natural selection [23,24]. This combination of accessibility, centralization, evolvability, and ability to produce very diverse outcomes is unparalleled and holds the promise of unique societal impact.

Conversely, distinctive hurdles involve automation capabilities that are nascent compared with other fields, and biology curricula that do not currently emphasize the backgrounds in mathematics and robotics that are critical for creating SDLs (see section '*Gaps for realizing self-driving labs*' for further discussion).

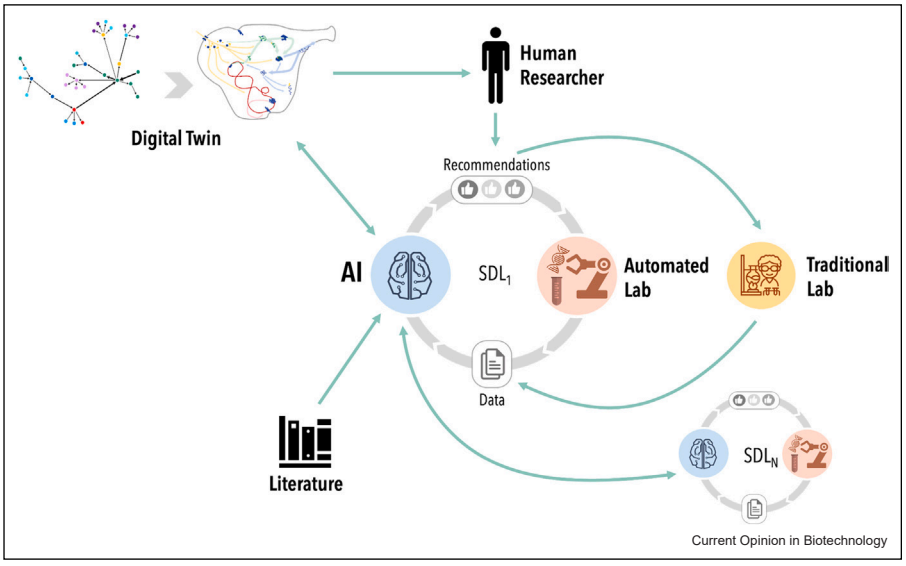
Benefits of self-driving labs

The main appeal for SDLs is their ability to enable significant scientific advances, which justifies their significant cost. These scientific advances involve, first, solving difficult biology questions that are intractable with current approaches. Second, and arguably more importantly, they involve upending the development of science as we know it, to accelerate it by leveraging AI.

The high level of investment needed to enable biological SDLs is only warranted if directed toward solving important and difficult biological problems. These involve biological problems that could take decades, or even centuries, to solve otherwise: for example, the prediction of protein structure from sequence [25]. Some examples of remaining difficult biological problems, including both topics of fundamental and practical importance, are:

1. *Systematic increase of Titer, Rate, and Yield (TRY) for bioengineered microbial strains*. A significant obstacle in developing commercially viable processes is reaching economically viable levels of TRY of a biologically produced small molecule. The traditional approach involves heuristic combinatorial processes that rely on

Figure 3



We envision SDLs working in a network of other SDLs and humans. To start with, in order for the SDL to make progress with respect to the current state of scientific knowledge, it must be able to draw information from existing literature. The SDL should also be able to communicate with other SDLs so as to efficiently partition the scientific phase space (i.e. the abstract space conformed by all possible configurable experiment parameter choices) to be explored (e.g. SDL_1 focuses on one promoter set and SDL_2 focuses on another promoter set). Current technical limitations limit our ability to automate all experiments, so SDLs should be able to produce unequivocal instructions for humans to follow in traditional labs, and ingest the data so produced. The final result of the operation of this network will be a digital twin of the system under study. This digital twin will likely start as a very crude and qualitative description of system parts and their connections, which will evolve as new information is obtained into more sophisticated mechanistic, quantitatively predictive models of the system under study (à la whole-cell model [28], sporting accurate predictions). These digital twins would be used by humans to access the scientific knowledge generated through this hybrid network and suggest their own recommendations. Whole-cell model figure adapted from Kerr et al. [28].

strain-specific in-depth metabolic knowledge (i.e. the'pull-push-block' approach [26]), and do not transfer well to other products, pathways, and hosts.

2. *Mapping of regulatory networks.* Perhaps the largest hurdle in predicting an organism's metabolism is to understand how it is regulated, which involves the mechanistic understanding of a large part of its genomic complement [27].
3. *Elucidating the genotype-to-phenotype link.* This challenge is, arguably, the central problem in biology, but despite promising advances [28–30], it remains beyond our reach to predict accurately and quantitatively the behavior of an organism given its genome.
4. *Inverse design of microbiomes.* Microbial communities exhibit remarkable capabilities, from driving Earth's biogeochemical cycles to increasing crop productivity [31]. However, we currently lack the knowledge to design communities to meet a specification: for example, remain stable over a year, or remove *X* grams/liter/hour of phosphorus from wastewater.
5. *Exploring biological behavior outside Earth.* Understanding how biological systems react to being in deep space or on another planet/satellite is fundamental to enable space exploration, and the proliferation of humankind beyond a single planet. However, workforce and equipment are extremely limited in space, due to the very high logistic cost of transporting them to orbit and beyond [32].

Each of these challenges will require very different robotic setups for the corresponding SDL. The cost of each of these SDLs would be directly related to its scope: SDLs exploring a large phase space and using sophisticated assays are bound to be costly, whereas simpler SDLs can potentially be quite affordable.

Perhaps, the most important impact of SDLs in science would come from the ability to automatically build scientific knowledge. By scientific knowledge, we mean a generalized body of facts, laws, and theories able to explain and predict the behavior of the system under study. We envision SDLs to be able to draw from prior knowledge and external sources as needed to perform experiments that improve this knowledge (Fig. 3). This improvement would be reflected in increased mechanistic understanding and predictive power. We envisage that a SDL would store its accumulating knowledge as a digital twin, whose role evolves as more is learned about the biological system it is analyzing. Digital twins are virtual replicas of real-world products, systems, beings, communities, or even cities, and have become critical assets for industry [33]. The initial role, in many cases, of

the digital twin would be simply to suggest experiments that identify the parts and their associations. Once sufficient experimental data have been generated to identify these associations, the role of the digital twin would be to suggest experiments that determine which correlations are causal. Once causal effects have been elucidated, the role of the digital twin would evolve into designing experiments to validate a mechanistic theory capable of explaining and quantitatively predicting these causal effects. Once this theory is calibrated, the role of the digital twin would transition into designing experiments that enable new biological systems to be built to a desired specification (inverse design). In the words of Feynman, "What I cannot create, I do not understand". We anticipate that this strategy to build scientific knowledge would involve a hybrid approach, combining pure SDLs with humans, traditional labs, and existing literature (Fig. 3).

Admittedly, this type of AI technology is not yet available, despite significant recent advances in question-answering and summarization [34], integrating prior knowledge into AI systems [35], and automated derivation of generalizable rules [36,37]. Massive language models such as GTP-3 are able to perform impressive tasks that appear to mimic natural language understanding, but these systems are ungrounded and are essentially performing pattern matching, and much needs to be done to unite classical symbolic reasoning systems with deep learning approaches [38]. Indeed, the scientific process of developing and experimentally testing hypotheses, to create a falsifiable worldview that can be used to make quantitative predictions and inform decision-making, comes quite close to the definition of artificial general intelligence.

Gaps for realizing self-driving labs

The benefits of SDLs necessitate several technological and social advances to become a reality. The gaps involve limitations in current automation technologies, AI algorithms, data management, and, importantly, sociological hurdles.

While automation of synthetic biology processes using liquid-handling commercial robotic workstations is gaining momentum, this approach has limitations for SDLs that new technologies may help ease. Companies such as, for example, Ginkgo Bioworks or Amyris automate their discovery process using these workstations, and a few are even providing automation as a service [39]. However, the processes automated in the chemistry and material sciences SDLs discussed above are only a subset of the ones needed in synthetic biology. Typical molecular biology processes such as cell transformations via electroporation, colony picking, plating, and outgrowth, while doable through liquid handlers and other instrumentation, are very difficult to link together in the seamless manner SDLs require (Fig. 2). Microfluidics offer the opportunity to provide this seamless integration by encapsulating cells and reagents into droplets, and manipulating them precisely. Indeed, microfluidic platforms have been proposed for miniaturization of biological reactions, including DNA synthesis and assembly [40], transformation [41,42], cell-free expression [43], and phenotypic screening by fluorescence [44] and mass spectrometry [45]. Truly disruptive functionalities can be achieved by combining these capabilities with new developments in molecular sensors embedded on semiconductor chips [46], wireless optically activated microscopic sensors [47], monitoring of free radicals through fluorescent nanodiamonds [48], metabolic modulation through optogenetics [49], or manipulation of cells with light [50]. Microfluidic sampling from bioreactors can also enable real-time sensing and imaging of cells in their environments, enabling continuous data capture. Moreover, these microfluidic platforms are far more affordable and use less reagents than robotic workstations, permitting a much larger number of experiments and democratizing the access to synthetic biology. Their routine use in synthetic biology, however, necessitates sustained investment to enable seamless functioning and the automation of the full range of synthetic biology processes.

Novel AI algorithms are needed to make SDLs a reality in synthetic biology. Although current algorithms can guide the metabolic engineering process effectively [51], widespread adoption of SDLs will require the AI to understand context, and the ability to produce interpretable knowledge. This means the ability to 1) use prior knowledge to inform the AI in the SDL, and 2) extract knowledge out of the predictive capabilities of the AI such that it can be extrapolated to related, but different, experimental conditions by other human researchers or SDLs (Fig. 3). The ability to leverage and produce extrapolatable knowledge is critical if we are to benefit from a large amount of SDLs. Otherwise, humans would become the bottleneck in transferring the knowledge accumulated in the digital twins from and to the SDLs (Fig. 3). One possibility to introduce this much-needed context may lie in the use of foundational models [52], trained on massive datasets, and adapted to specific use cases.

Data management is a critical link between automation and AI algorithms that has been often neglected in the past. While often considered a burdensome chore, there is simply no AI without data, and there are no SDLs without AI. General ontologies and extensible standards for data and protocols are critical if large amounts of data are to be collected and seamlessly integrated into an ecosystem involving continuous data exchange among SDLs and human researchers.

Another important obstacle for the creation of SDLs in biology involves the sociological challenges in having computer scientists and automation engineers work together with molecular and synthetic biologists. These two worlds embody very different scientific cultures, which are reflected not only in how they solve problems, but also which problems they consider worth solving [53]. Having them work together constructively is, arguably, harder than the technological challenges faced by SDLs in biology. Currently, computational and bench scientists are trained very differently: a critical first step is to design a training curriculum that exposes them to each other's world.

Conclusion

While SDLs are bound to be costly endeavors, the expected returns make them worthwhile undertakings. A fully functioning network of SDLs and human researchers (Fig. 3) would not only provide significant biological knowledge, but also the ability to fully exploit synthetic biology for biomanufacturing purposes. Furthermore, they would provide the opportunity to understand and improve the process of constructing scientific knowledge. In that sense, the large project of creating SDLs mirrors the Human Genome Project, in that they show a potential to fundamentally transform the field of biology.

We must, however, be aware of the risks associated with SDLs: their use for nefarious purposes (e.g. virus synthesis), including the ability to be manipulated via remote cyberattacks. A more subtle risk involves the possible long-term misalignment with our values and goals, which can be challenging to fully encode in a machine-readable manner, potentially allowing the system to act in an unintended or undesired manner.

Despite the risks and challenges, we believe that SDLs represent the next leap forward in the progress of scientific research, and that synthetic biology poses a unique opportunity for their development.

Conflict of interest statement

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests: Nathan Hillson has financial interests in TeselaGen Biotechnologies and Ansa Biotechnologies.

Data Availability

No data were used for the research described in the article.

Acknowledgements

This work was part of the DOE Agile BioFoundry (http://agilebiofoundry.org), the Advanced Biofuels and Bioproducts Process Development Unit (https://abpdu.lbl.gov/), and the DOE Joint BioEnergy Institute (http://www.jbei.org), supported by the U. S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, and the Office of Science, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. S.P. and D.A. were supported by Laboratory Directed Research and Development (LDRD) funds provided by Lawrence Berkeley National Laboratory, operated for the U.S. Department of Energy under the same contract. H.G.M. was also supported by the Basque Government through the BERC 2018–2021 program and by the Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2017-0718. K.E.B. was funded by the Department of Energy, Advanced Scientific Computing Research. J.M.M. was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory (LLNL) SFA "From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms," under Contract DE-AC52-07NA27344 (LLNL-JRNL-837127). J.M.C. was supported in part by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office under contract DE-EE0008927. Gy.B. was supported by the "Rapid Design and Engineering of Smart and Secure Microbiological Systems" project funded by the Biological Systems Science Division's Genomic Science Program, within the U.S. Department of Energy, Office of Science, Biological and Environmental Research. Argonne National Laboratory is managed by UChicago Argonne, LLC for DOE under contract number DE-AC02-06CH11357. LW was funded by the US National Science Foundation Graduate Research Fellowship. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility or the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this paper, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). Funding for open-access charge: US Department of Energy.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest.

1. Häse F, Roch LM, Aspuru-Guzik A: **Next-generation experimentation with self-driving laboratories**. *Trends Chem* 2019, **1**:282-291, https://doi.org/10.1016/j.trechm.2019.02.007

2. Soldatov MA, Butova VV, Pashkov D, Butakova MA, Medvedev PV, Chernov AV, *et al.*: **Self-driving laboratories for development of new functional materials and optimizing known reactions**. *Nanomater.* 2021, **11**:619, https://doi.org/10.3390/nano11030619

3. Bennett JA, Abolhasani M: **Autonomous chemical science and engineering enabled by self-driving laboratories**. *Curr Opin Chem Eng* 2022, **36**:100831, https://doi.org/10.1016/j.coche.2022.100831

4. Beal J, Rogers M: **Levels of autonomy in synthetic biology engineering**. *Mol Syst Biol* 2020, **16**:e10019, https://doi.org/10.15252/msb.202010019

5. Vrana J, de Lange O, Yang Y, Newman G, Saleem A, Miller A, *et al.*: **Aquarium: open-source laboratory software for design, execution and data management**. *Synth Biol* 2021, **6**:ysab006, https://doi.org/10.1093/synbio/ysab006

6. • King RD, Whelan KE, Jones FM, Reiser PGK, Bryant CH, Muggleton SH, *et al*.: **Functional genomic hypothesis generation and experimentation by a robot scientist**. *Nature* 2004, **427**:247-252, https://doi.org/10.1038/nature02236.

Adam is the first published example of a closed-loop system that designs and executes experiments to test inferred hypotheses. A classic well before SDLs became of widespread interest.

7. • Williams K, Bilsland E, Sparkes A, Aubrey W, Young M, Soldatova LN, *et al*.: **Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases**. *J R Soc Interface* 2015, **12**:20141289, https://doi.org/10.1098/rsif.2014.1289.

Eve constitutes an outstanding example of the use of SDLs to alleviate the large cost of drug discovery.

8. •• Granda JM, Donina L, Dragone V, Long D-L, Cronin L: **Controlling an organic synthesis robot with machine learning to search for new reactivity**. *Nature* 2018, **559**:377-381, https://doi.org/10.1038/s41586-018-0307-8.

A great example of the potential of SDLs, showing how a robot is able to systematically explore chemical space and successfully predict reactivity.

9. Christensen M, Yunker LPE, Adedeji F, Häse F, Roch LM, Gensch T, *et al*.: **Data-science driven autonomous process optimization**. *Commun Chem* 2021, **4**:112, https://doi.org/10.1038/s42004-021-00550-x

10. Wang L, Karadaghi LR, Brutchey RL, Malmstadt N: **Self-optimizing parallel millifluidic reactor for scaling nanoparticle synthesis**. *Chem Commun* 2020, **56**:3745-3748, https://doi.org/10.1039/d0cc00064g

11. MacLeod BP, Parlane FGL, Morrissey TD, Häse F, Roch LM, Dettelbach KE, *et al*.: **Self-driving laboratory for accelerated discovery of thin-film materials**. *Sci Adv* 2020, **6**:eaaz8867, https://doi.org/10.1126/sciadv.aaz8867

12. MacLeod BP, Parlane FGL, Rupnow CC, Dettelbach KE, Elliott MS, Morrissey TD, *et al*.: **A self-driving laboratory advances the Pareto front for material properties**. *Nat Commun* 2022, **13**:995, https://doi.org/10.1038/s41467-022-28580-6

13. Shimizu R, Kobayashi S, Watanabe Y, Ando Y, Hitosugi T: **Autonomous materials synthesis by machine learning and robotics**. *APL Mater* 2020, **8**:111110, https://doi.org/10.1063/5.0020370

14. Gongora AE, Xu B, Perry W, Okoye C, Riley P, Reyes KG, *et al*.: **A Bayesian experimental autonomous researcher for mechanical design**. *Sci Adv* 2020, **6**:eaaz1708, https://doi.org/10.1126/sciadv.aaz1708

15. Rooney MB, MacLeod BP, Oldford R, Thompson ZJ, White KL, Tungjunyatham J, *et al*.: **A self-driving laboratory designed to accelerate the discovery of adhesive materials**. *Digit Discov* 2022 **1**:382-389, https://doi.org/10.1039/D2DD00029F

16. Deneault JR, Chang J, Myung J, Hooper D, Armstrong A, Pitt M, *et al*.: **Toward autonomous additive manufacturing: Bayesian optimization on a 3D printer**. *MRS Bull* 2021, **46**:566-575, https://doi.org/10.1557/s43577-021-00051-1

17. Dave A, Mitchell J, Kandasamy K, Wang H, Burke S, Paria B, *et al*.: **Autonomous discovery of battery electrolytes with robotic experimentation and machine learning**. *Cell Rep Phys Sci* 2020, **1**:100264, https://doi.org/10.1016/j.xcrp.2020.100264

18. •• Burger B, Maffettone PM, Gusev VV, Aitchison CM, Bai Y, Wang X, *et al*.: **A mobile robotic chemist**. *Nature* 2020, **583**:237-241, https://doi.org/10.1038/s41586-020-2442-2.

An inspiring use of a mobile robotic arm to automate the researcher rather than the instruments, opening the transition to SDLs for any traditional lab.

19. Si T, Chao R, Min Y, Wu Y, Ren W, Zhao H: **Automated multiplex genome-scale engineering in yeast**. *Nat Commun* 2017, **8**:15187, https://doi.org/10.1038/ncomms15187

20. HamediRad M, Chao R, Weisberg S, Lian J, Sinha S, Zhao H: **Towards a fully automated algorithm driven platform for biosystems design**. *Nat Commun* 2019, **10**:5150, https://doi.org/10.1038/s41467-019-13189-z

21. Kanda GN, Tsuzuki T, Terada M, Sakai N, Motozawa N, Masuda T, *et al*.: **Robotic search for optimal cell culture in regenerative medicine**. *eLife* 2022, **11**:e77007, https://doi.org/10.7554/eLife.77007

22. Knott GJ, Doudna JA: **CRISPR-Cas guides the future of genetic engineering**. *Science* 2018, **361**:866-869, https://doi.org/10.1126/science.aat5011

23. Zhong Z, Wong BG, Ravikumar A, Arzumanyan GA, Khalil AS, Liu CC: **Automated continuous evolution of proteins in vivo**. *ACS Synth Biol* 2020, **9**:1270-1276, https://doi.org/10.1021/acssynbio.0c00135

24. Javanpour AA, Liu CC: **Evolving small-molecule biosensors with improved performance and reprogrammed ligand preference using OrthoRep**. *ACS Synth Biol* 2021, **10**:2705-2714, https://doi.org/10.1021/acssynbio.1c00316

25. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al*.: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583-589, https://doi.org/10.1038/s41586-021-03819-2

26. Yan Q, Pfleger BF: **Revisiting metabolic engineering strategies for microbial synthesis of oleochemicals**. *Metab Eng* 2020, **58**:35-46, https://doi.org/10.1016/j.ymben.2019.04.009

27. Herrgård MJ, Covert MW, Palsson BØ: **Reconstruction of microbial transcriptional regulatory networks**. *Curr Opin Biotechnol* 2004, **15**:70-77, https://doi.org/10.1016/j.copbio.2003.11.002

28. • Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, *et al*.: **A whole-cell computational model predicts phenotype from genotype**. *Cell* 2012, **150**:389-401, https://doi.org/10.1016/j.cell.2012.05.044.

One of the first examples of a whole-cell model, accounting for all annotated gene functions in *Mycoplasma genitalium*, and validated against a broad range of data.

29. Macklin DN, Ahn-Horst TA, Choi H, Ruggero NA, Carrera J, Mason JC, *et al*.: **Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation**. *Science* 2020, **369 (6502)**, https://doi.org/10.1126/science.aav3751

30. Thornburg ZR, Bianchi DM, Brier TA, Gilbert BR, Earnest TM, Melo MCR, *et al*.: **Fundamental behaviors emerge from simulations of a living minimal cell**. *Cell* 2022, **185**:345-360.e28, https://doi.org/10.1016/j.cell.2021.12.025

31. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE, O'Malley MA, *et al*.: **Common principles and best practices for engineering microbiomes**. *Nat Rev Microbiol* 2019, **17**:725-741, https://doi.org/10.1038/s41579-019-0255-9

32. Lauren, Yang J, Scott R, Qutub A, Martin H, Berrios D, et al.: **Beyond Low Earth Orbit: Biological Research, Artificial Intelligence, and Self-Driving Labs**.

33. • Jiang Y, Yin S, Li K, Luo H, Kaynak O: **Industrial applications of digital twins**. *Philos Trans A Math Phys Eng Sci* 2021, **379**:20200360, https://doi.org/10.1098/rsta.2020.0360.

Good introduction to digital twins, and how they are becoming an industry staple.

34. Neves M, Leser U: **Question answering for biology**. *Methods* 2015, **74**:36-46, https://doi.org/10.1016/j.ymeth.2014.10.023

35. •• Cai S, Mao Z, Wang Z, Yin M, Karniadakis GE: **Physics-informed neural networks (PINNs) for fluid mechanics: a review**. *Acta Mech Sin* 2021, **37**:1727-1738, https://doi.org/10.1007/s10409-021-01148-1.

An informative review on how to embed prior knowledge in AI, in this case for fluid dynamics in the form of PINNs (physics-informed neural networks). Similar approaches would be needed for biology.

36. Liu Z, Tegmark M: **Machine learning conservation laws from trajectories**. *Phys Rev Lett* 2021, **126**:180604, https://doi.org/10.1103/PhysRevLett.126.180604

37. •• Guimerà R, Reichardt I, Aguilar-Mogas A, Massucci FA, Miranda M, Pallarès J, *et al*.: **A Bayesian machine scientist to aid in the solution of challenging scientific problems**. *Sci Adv* 2020, **6**:eaav6971, https://doi.org/10.1126/sciadv.aav6971.

A stimulating demonstration of the power of 'machine scientists', able to extract closed mathematical models automatically out of data.

38. d'Avila GA, Lamb LC: **Neurosymbolic AI: the 3rd wave**. *arXiv* 2020, https://doi.org/10.48550/arxiv.2012.05876

39. Arnold C: **Cloud labs: where robots do the research**. *Nature* 2022, **606**:612-613, https://doi.org/10.1038/d41586-022-01618-x

40. Lee C-C, Snyder TM, Quake SR: **A microfluidic oligonucleotide synthesizer**. *Nucleic Acids Res* 2010, **38**:2514-2521, https://doi.org/10.1093/nar/gkq092

41. • Gach PC, Shih SCC, Sustarich J, Keasling JD, Hillson NJ, Adams PD, *et al*.: **A droplet microfluidic platform for automating genetic engineering**. *ACS Synth Biol* 2016, **5**:426-433, https://doi.org/10.1021/acssynbio.6b00011.

A nice demonstration of what microfluidics can achieve in terms of automating synthetic biology protocols.

42. Iwai K, Wehrs M, Garber M, Sustarich J, Washburn L, Costello Z, *et al*.: **Scalable and automated CRISPR-based strain engineering using droplet microfluidics**. *Micro Nanoeng* 2022, **8**:31, https://doi.org/10.1038/s41378-022-00357-3

43. Hori Y, Kantak C, Murray RM, Abate AR: **Cell-free extract based optimization of biomolecular circuits with droplet microfluidics**. *Lab Chip* 2017, **17**:3037-3042, https://doi.org/10.1039/c7lc00552k

44. Iwai K, Ando D, Kim PW, Gach PC, Raje M, Duncomb TA, et al.: **Automated flow-based/digital microfluidic platform integrated with onsite electroporation process for multiplex genetic engineering applications**. In *Proceedings of the 2018 IEEE Micro Electro Mechanical Systems (MEMS)*. IEEE; 2018:1229–1232. doi:10.1109/MEMSYS.2018.8346785.

45. Heinemann J, Deng K, Shih SCC, Gao J, Adams PD, Singh AK, *et al*.: **On-chip integration of droplet microfluidics and nanostructure-initiator mass spectrometry for enzyme screening**. *Lab Chip* 2017, **17**:323-331, https://doi.org/10.1039/c6lc01182a

46. •• Fuller CW, Padayatti PS, Abderrahim H, Adamiak L, Alagar N, Ananthapadmanabhan N, *et al*.: **Molecular electronics sensors on a scalable semiconductor chip: a platform for single-molecule measurement of binding kinetics and enzyme activity**. *Proc Natl Acad Sci USA* (5) 2022, **119**:e2112812119, https://doi.org/10.1073/pnas.2112812119.

A very interesting report on the possibilities created by embedding single molecules in electronic chips.

47. Cortese AJ, Smart CL, Wang T, Reynolds MF, Norris SL, Ji Y, *et al*.: **Microscopic sensors using optical wireless integrated circuits**. *Proc Natl Acad Sci USA* 2020, **117**:9173-9179, https://doi.org/10.1073/pnas.1919677117

48. Nie L, Nusantara AC, Damle VG, Sharmin R, Evans EPP, Hemelaar SR, *et al*.: **Quantum monitoring of cellular metabolic activities in single mitochondria**. *Sci Adv* (21) 2021, **7**:eabf0573, https://doi.org/10.1126/sciadv.abf0573

49. Wegner SA, Barocio-Galindo RM, Avalos JL: **The bright frontiers of microbial metabolic optogenetics**. *Curr Opin Chem Biol* 2022, **71**:102207, https://doi.org/10.1016/j.cbpa.2022.102207

50. •• Rienzo M, Lin K-C, Mobilia KC, Sackmann EK, Kurz V, Navidi AH, *et al*.: **High-throughput optofluidic screening for improved microbial cell factories via real-time micron-scale productivity monitoring**. *Lab Chip* 2021, **21**:2901-2912, https://doi.org/10.1039/d1lc00389e.

This paper demonstrates the use of microfluidics and automated cell manipulation through light for synthetic biology, providing a promising platform for SDLs.

51. • Lawson CE, Martí JM, Radivojevic T, Jonnalagadda SVR, Gentz R, Hillson NJ, *et al*.: **Machine learning for metabolic engineering: a review**. *Metab Eng* 2021, **63**:34-60, https://doi.org/10.1016/j.ymben.2020.10.005.

Interesting review on the current and possible applications of AI in metabolic engineering and synthetic biology.

52. **On the Opportunities and Risks of Foundation Models**; 2021 [cited 15 Aug 2022]. Available from: ⟨https://fsi.stanford.edu/publication/opportunities-and-risks-foundation-models⟩.

53. Eslami M, Adler A, Caceres RS, Dunn JG, Kelley-Loughnane N, Varaljay VA, *et al*.: **Artificial intelligence for synthetic biology**. *Commun ACM* 2022, **65**:88-97, https://doi.org/10.1145/3500922