


# In silico design and automated learning to boost next-generation smart biomanufacturing

Pablo Carbonell <sup>1,2,\*</sup>, Rosalind Le Feuvre<sup>1,\*</sup>, Eriko Takano<sup>1</sup>, and Nigel S. Scrutton<sup>1</sup>

<sup>1</sup>Manchester Synthetic Biology Research Centre for Fine and Speciality Chemicals (SYNBIOCHEM) and Future Biomanufacturing Research Hub, Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, UK and <sup>2</sup>Instituto Universitario de Automática e Informática Industrial, Universitat Politècnica de València, 46022 Valencia, Spain

\*Corresponding authors: E-mails: pablo.carbonell@manchester.ac.uk and r.a.le-feuvre@manchester.ac.uk

## Abstract

The increasing demand for bio-based compounds produced from waste or sustainable sources is driving biofoundries to deliver a new generation of prototyping biomanufacturing platforms. Integration and automation of the design, build, test and learn (DBTL) steps in centers like SYNBIOCHEM in Manchester and across the globe (Global Biofoundries Alliance) are helping to reduce the delivery time from initial strain screening and prototyping towards industrial production. Notably, a portfolio of producer strains for a suite of material monomers was recently developed, some approaching industrial titers, in a *tour de force* by the Manchester Centre that was achieved in less than 90 days. New *in silico* design tools are providing significant contributions to the front end of the DBTL pipelines. At the same time, the far-reaching initiatives of modern biofoundries are generating a large amount of high-dimensional data and knowledge that can be integrated through automated learning to expedite the DBTL cycle. In this Perspective, the new design tools and the role of the learning component as an enabling technology for the next generation of automated biofoundries are discussed. Future biofoundries will operate under completely automated DBTL cycles driven by *in silico* optimal experimental planning, full biomanufacturing devices connectivity, virtualization platforms and cloud-based design. The automated generation of robotic build worklists and the integration of machine-learning algorithms will collectively allow high levels of adaptability and rapid design changes toward fully automated smart biomanufacturing.

**Key words:** synthetic biology; biomanufacturing; automation; machine learning

## 1. Next-generation smart biomanufacturing

As the global demand for bio-based products and materials steadily increases (1), a new generation of automated smart biomanufacturing laboratories is transforming the industry. Traditionally, metabolic engineering projects have been done manually, often involving trial-and-error. Today, biomanufacturing is moving to replace some of the chemical industry processes through more sustainable bio-based solutions for production not only of fine and specialty chemicals but also of

commodity compounds. Biofoundries are facilities for engineering biology that aim to make such processes more efficient, systematic and standardized through the application of the Design/Build/Test/Learn cycle (2), similar to other manufacturing industries. Process automation is required in order to reach competitive time frames from initial strain screening and prototyping to scale up. Automated design for biomanufacturing is a central piece of such an approach that takes place in the initial steps of the biomanufacturing project (Figure 1). Whilst automated design has received some interest from the community

Submitted: 29 May 2020; Received (in revised form): 8 September 2020; Accepted: 28 September 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

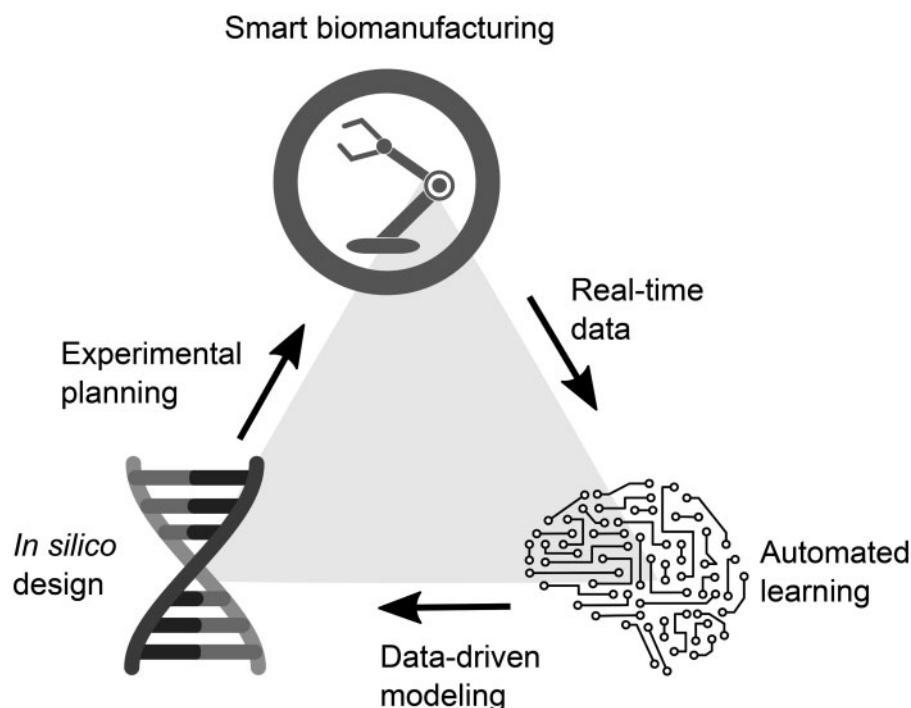


Figure 1. *In silico* design driven by automated learning from real-time data directs experimental planning for smart biomanufacturing.

(3), it is still not fully integrated due to the breadth and complexity of biological design (4), and the lack of appropriate standards for data exchange and standardized input and output formats (5). At the other end of the design, build, test and learn (DBTL) cycle, automated learning has been hindered so far because of the lack of rich datasets, i.e., big data, for training the models and the lack of interlab data in the cloud repositories allowing the gathering of information from diverse contexts (6). Here, we discuss how ongoing integration of automated design and learning will boost future biomanufacturing technologies, allowing rapid on-demand production of bespoke bio-based chemicals in smart biomanufacturing foundries.

Smart manufacturing is a technology that responds in real-time to meet changing demands (7). Biofoundries operating under the principle of smart biomanufacturing should meet the changes in the supply network for raw feedstocks as well as for synthetic genetic parts from DNA foundries; respond agilely to customers changing needs; and ensure productivity, quality, delivery and flexibility. However, meeting such demands is a challenging endeavor because of the reproducibility issues that have plagued bioproduction technologies (8) and because of the often low performance of fermentation processes in terms of titer, rates and yield and across different scales and process modes (9). Such issues are now being addressed through laboratory automation, standardization of protocols and communication between equipment. Using sensors and wireless technologies (10), cloud laboratories are being established worldwide (8) and production lines can be shared between biofoundries. Notably, the Global Biofoundries Alliance (11) is an initiative that is facilitating the collaboration between partners across the world. Several working groups are advancing towards better standardization and interoperability of software for controlling production processes and resource planning.

Automated optimization of the processes is becoming possible through robotic fermentation platforms, where growing conditions can be highly monitored and controlled in a parallelized

fashion with early consideration of scale up (12). The wealth of big data that current omics techniques provide allow the real-time analysis of proteomics, transcriptomics and metabolomics data. However, high-throughput screening can be a limitation for certain classes of molecules, for instance, if no biosensor exists. Powerful machine-learning algorithms in combination with automation can then be used to exploit the data and predict new interventions for process optimization (13). In recent years, the number of applications of machine learning into biotechnological processes has notably increased thanks to a new generation of algorithms such as deep learning as well as the deployment of easy-to-use development frameworks (14). Deep learning algorithms are a family of neural networks methods where layers are generalized and organized in complex architectures such as convolutional networks. For instance, deep learning was applied by Wu *et al.* (15) to assist the directed evolution of an enzyme in order to produce the two enantiomers of a new-to-nature carbene Si-H insertion reaction. Sampling the combinatorial sequence space by directed evolution alone would have been too expensive, but machine-learning models trained on tested variants provided a faster method by allowing its computational exploration (16). Similarly, deep learning has been applied to biological sequence data (10) in order to perform functional prediction. Amidi *et al.* (17) used 3D convolutional networks, a class of deep learning algorithms, to classify enzymes by their EC class in the entire PDB database based on spatial structure. Machine learning has been applied to different strategies for metabolic engineering (18), for instance, Costello and Martin (19) combined machine learning with multiomics data (proteomics and metabolomics) to predict the kinetic model of engineered metabolic pathways. In yet another example, Oopenorth *et al.* (20) used the data produced in the first Design/Build/Test/Learn cycle to train several machine-learning algorithms and to suggest protein profiles for a second cycle that would increase production. Similarly, the cell-free iPROBE platform for prototyping and rapid optimization (21)

implemented a neural network to predict beneficial combinations of biosynthetic enzymes in a pathway. Several examples also exist of the application of machine learning to synthetic biology, such as in the prediction of success of rapid synthesis of long DNA fragments (22) the lab-of-origin of the engineered DNA (12), or in the design of simultaneously stable single-guide RNAs for CRISPR interference (23), context-aware DNA design in translational regulation for limonene production in *Escherichia coli* (24), as well as in other strategies for the design of gene circuits (25).

## 2. *In silico* design tools for biomanufacturing

As a major contributor to the new fully automated pipeline paradigm, the concerted orchestration of a growing ecosystem of *in silico* design tools is undeniably transforming metabolic engineering into a biomanufacturing technology. For some years design tools were considered as simply assistants for individual tasks, but the current view is that integrated *in silico* tools should underpin the automated pipeline for the design stage of smart biomanufacturing. The first task of metabolic pathway design automation is the selection of chemical targets. Generally, targets are imposed by external partners such as customers interested in commercializing the bio-based version of some compound of interest like a natural product, a pharmaceutical or a material building block, to name a few. Another consideration in the development of sustainable industrial systems is the use of sustainable feedstocks or utilization of waste products in a bid to reduce resource consumption (including reliance on precious metals and decreasing petrochemical consumption) and greenhouse gas emissions, this is increasingly important in the bid to mitigate climate change and ensure clean growth (26). Recently, several proposals have appeared trying to rationalize automating the target selection in the same fashion as in the chemical industry. Life cycle and techno-economic analyses (27) are critical steps used to assess the potential economic viability of production by microbial fermentation of the target from raw materials at an industrial plant scale, as well as assessing the value of new chemical production routes that circumvent existing intellectual property (28). Whilst these are at present mostly performed manually they show potential for achieving automated target selection (29).

Moreover, the bio-based production of chemical targets requires the identification of viable routes. It is possible that a natural biosynthetic pathway exists, but it might have been optimized through evolution in a different host (30). Therefore, even when routes are available, it is often preferable to select enzymes from different sources. A more ambitious effort consists of looking for alternative pathways other than the natural one or the engineered ones reported in the literature. To that end, retrosynthetic analysis is performed (31). Several computational platforms for retrobiosynthesis exist (32, 33), such as BNICE (33), GEM-Path (34), RetroPath 2.0 (35), PathPred (36) or Cho *et al.* (37). Each one differs from the others on their unique computational representation of reactions and the way predicted pathways are ranked. However, there is a sizable gap between the virtually huge number of alternative routes that can be computationally predicted and the few that because of limitations in the total number of experimental runs to be performed would finally be tested in the laboratory. Because of this challenge, retrosynthetic algorithms have been increasingly refined with the hope of becoming ‘smarter’ and therefore capable of providing routes that are the closest to the expert

suggestions. Using state-of-the-art machine-learning algorithms such as deep learning, it has been shown recently that automated retrosynthetic analysis for synthetic chemistry can beat the experts (38). Retrosynthetic analysis for biosynthesis, i.e. bioretrosynthesis, is more challenging because it needs to combine chemical reaction knowledge with enzyme biocatalysis expertise. Despite this complexity, recent progress using reaction rules with associated scores is helping in narrowing down the search to the most promising routes (39), as exemplified in RetroPath2.0 (35).

When performing pathway selection, it would be of little use to select one of the predicted pathways through retrosynthesis if enzymes with the required efficiency cannot be found for individual steps of the pathway. Enzyme selection can be done based on homology, i.e. by performing a BLAST search of homologs, as performed by EC-Blast (40) and then selecting them either because of their reported activity in databases like BRENDA or any diversification approach. Moreover, the search can be complemented by looking for reactions that are similar to the target reaction. This is the strategy developed by Selenzyme (41), which is well-fitted for annotating retrosynthetic searches that are based on reaction rules. This tool provides key information about enzymes based on several criteria such as a reaction and sequence similarity; phylogenetic distance between source organism and host strain; sequence conservation; and other relevant properties. Other sources of information for the enzyme selection include structural analysis of the ensemble of homologs and prediction and identification of active sites and beneficial mutations. Enzymes can be screened for efficiency, and directed evolution (42) is generally then used in order to evolve variants with improved efficiency with respect to the native counterpart by defining the set of amino acid regions in the enzyme sequence to be mutated in order to improve its functionality. Such a process can be automated by using technologies like deep learning (15, 43). Several methods based on deep learning have been proposed for enzyme sequence design, for instance, to classify enzymes by EC number based on sequence (44, 45) or in enzyme engineering (46).

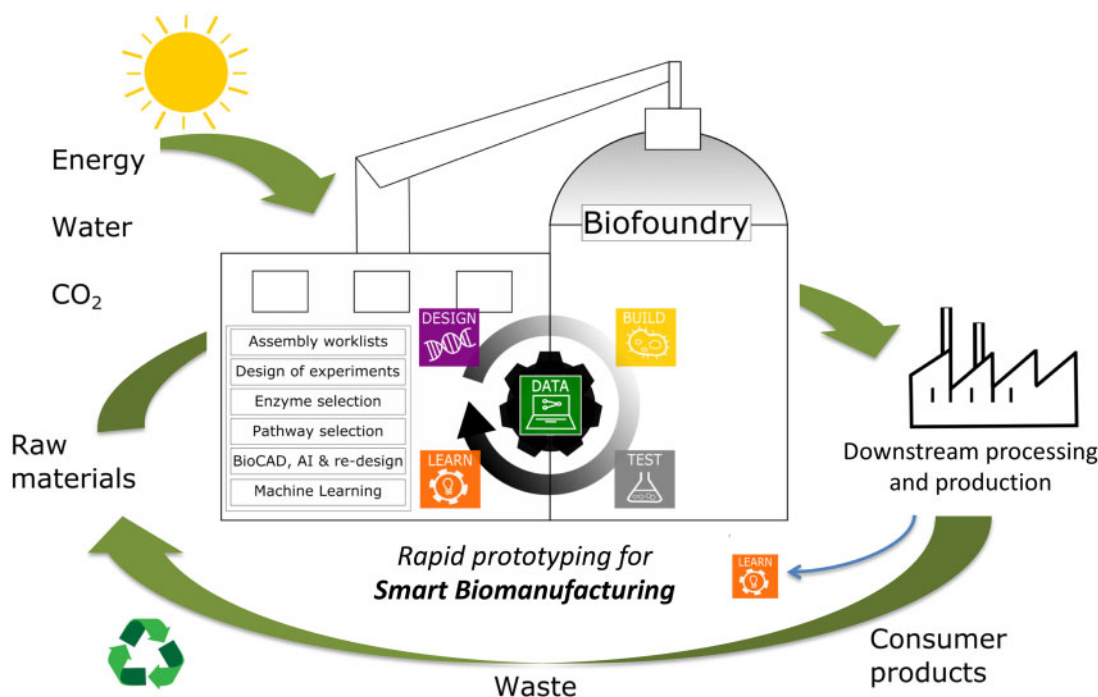
*In silico* enzyme selection is generally followed by genetic parts selection, both at transcription and translation levels. Transcriptional selection typically involves promoter design (47) using tools such as SelProm (48) or ePathOptimize (49), which has been applied to the violacein pathway. Translation design tools like the RBS Calculator (50) and RBSDesigner (51) are facilitating the bottom-up genetic circuit design. Tools are also guiding the design of smart combinatorial genetic circuit libraries like RedLibs (52) and PartsGenie (53) as well as automated circuit design such as include Cello (3), which can automatically compute the circuit implementing the desired logic function. All of these tools combined together with the increasingly more efficient high-throughput capabilities of smart biomanufacturing platforms, are nowadays optimized by the design of experiments (DoE) approaches (54). Through the optimal experimental design, smart biomanufacturing platforms can rapidly develop high producer titers based on a reduced set of experimental runs. The optimal design will generally follow the strategy of assuming simple models and relationships between the factors for rapid prototyping and identification of main factors among gene variants, promoters and other genetic parts. Those main identified factors can be then explored in depth through focused combinatorial libraries in order to model complex interactions through machine-learning algorithms.

### 3. The future of biomanufacturing is smart

Microbial strain engineering for on-demand production of chemicals is a major focus of biomanufacturing foundries that are developing automated pipelines that are agnostic to the target compound (2). Ultimately, the suite of fully automated design protocols would be integrated in the pipelines to allow seamless transition from chemical target to automated generation of build workflows. In addition, the development of cloud-based platforms would allow the remote implementation of design stages that feed into distributed (or distant) wet-lab facilities with experimental data automatically feeding back into learning steps for optimized re-design and iterative cycles for optimization (8, 55). Broadening the portfolio of design tools and providing access will allow the sharing of capabilities and the development of virtual *in silico* design. However, the wide uptake of design platforms will depend on the standardization of biological engineering and ensuring that these tools are accessible to the non-expert user.

Towards that end, several initiatives have been proposed in order to define metrics and assess the capabilities of biofoundries. In one of such initiative, a timed 'pressure test' administered by the US Department of Advanced Research Projects Agency (DARPA) consisting on building organisms to produce 10 molecules ran for 3 months in order to assess the foundry (56). Interestingly, several approaches were applied including retrosynthetic design and cell-free systems, achieving an outcome of 6/10 targets produced during the performance period. As a recent milestone, the Manchester SYNBIOCHEM Centre focused its DBTL pipeline on the rapid prototyping of microbial factories for the biosynthetic production of monomers for enhanced biomanufacturing of materials (57). The *in silico* design phase from initial target and pathway selection through to the design using tools such as RetroPath2.0, Selenzyme and PartsGenie and

ordering of DNA parts took ~10 days. Over 85 days, more than 160 genetic parts were designed and assembled into 115 unique biosynthetic pathway constructs and tested for *in vivo* production in *E. coli*. Targeting chemically diverse industrially relevant materials building blocks, the Centre's rapid prototyping capability successfully produced 17 materials monomers at competitive titers and key intermediates to demonstrate the potential of biofoundries in leading the sustainable production of next-generation materials. The ability to produce strains for gateway chemicals (e.g. flavanones (58)) with broad applications such as precursors to materials and active pharmaceutical ingredients, will be a major step forward. One such materials monomer targeted was enantioselective production of mandelic acid, a monomer for degradable thermoplastics with polystyrene-like properties (59) that can also serve in its enantiopure form as a building block for the synthesis of pharmaceuticals (60), with rapid scale-up delivery of g/l quantities in bioreactor cultures achieved. Whilst the potential for biobased production of key compounds has been demonstrated, in order for bio-based manufacturing to become a reality, the whole ecosystem of production needs to be considered with substantial reductions in the costs of production to become commercially viable (Figure 2). Important considerations include both upstream and downstream processes including the development of new more environmental friendly solvent/extractant systems with reduced energy consumption and loss (61, 62); greener product recovery (63) and separation processes (64); choice and cost of feedstocks and substrates; and the potential to utilize low/zero-cost waste streams (e.g. lignocellulose and glycerol waste from biodiesel production), or CO<sub>2</sub> (65–67). The engineering of robust industrial microbial strains for chemicals production at scale and to reduce the economic burden of biomanufacturing will require systems level design and implementation of integrated



**Figure 2.** Biofoundry model for rapid prototyping and smart biomanufacturing based on the automated Design, Build, Test and Learn (DBTL) cycle integrated with high-dimensional data and the principles of the circular economy. A suite of *in silico* machine learning-driven BioCAD tools for parts selection, build assembly and experimental test can leverage the agile circular process for downstream processing (DSP) and production of consumer products in order to achieve sustainable solutions that utilize renewable/waste feedstocks and replace the use of fossil fuels.



bioproduction platforms (56). Full techno-economic assessment will also be important to consider the commercial viability including production modes, separation technology and the price point of target molecules (e.g. commodity vs high value). Standardization, robotics and automation which will play an important part (8) in the development of ‘smart biomanufacturing’, are the focus of many research groups around the world, including the Global Biofoundry Alliance (<https://biofoundries.org>). Finally, coupling to ‘greener’ downstream processing (e.g. (68)) and waste recycling will be essential for the successful delivery of a circular biobased economy.

## Acknowledgements

P.C. acknowledges support from the Universitat Politècnica de València Talento Programme.

## Funding

The Biotechnology and Biological Sciences Research Council (BBSRC) and the Engineering and Physical Sciences Research Council (EPSRC) under grants: ‘Centre for synthetic biology of fine and specialty chemicals (SYNBIOCHEM)’ (BB/M017702/1) and ‘Future Biomanufacturing Research Hub’ (EP/S01778X/1). This work also received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement number 814408 (ShikiFactory100) and from MINECO/AEI EU grant number DPI2017-82896-C2-2-R.

*Conflict of interest statement.* None declared.

## References

- Le Feuvre, R.A. and Scrutton, N.S. (2018) A living foundry for Synthetic Biological Materials: a synthetic biology roadmap to new advanced materials. *Synth. Syst. Biotechnol.*, 3, 105–112.
- Carbonell, P., Jervis, A.J., Robinson, C.J., Yan, C., Dunstan, M., Swainston, N., Vinaixa, M., Hollywood, K.A., Currin, A., Rattray, N.J.W. et al. (2018) An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol.*, 1, 66.
- Nielsen, A.A.K., Der, B.S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E.A., Ross, D., Densmore, D. and Voigt, C.A. (2016) Genetic circuit design automation. *Science* (80-), 352, aac7341–aac7341.
- Kwok, R. (2010) Five hard truths for synthetic biology. *Nature*, 463, 288–290.
- Kitney, R., Adeogun, M., Fujishima, Y., Goñi-Moreno, Á., Johnson, R., Maxon, M., Steedman, S., Ward, S., Winickoff, D. and Philp, J. (2019) Enabling the advanced bioeconomy through public policy supporting biofoundries and engineering biology. *Trends Biotechnol.*, 37, 917–920.
- El Karoui, M., Hoyos-Flight, M. and Fletcher, L. (2019) Future trends in synthetic biology—a report. *Front. Bioeng. Biotechnol.*, 7, 175.
- Kang, H.S., Lee, J.Y., Choi, S., Kim, H., Park, J.H., Son, J.Y., Kim, B.H. and Noh, S.D. (2016) Smart manufacturing: past research, present findings, and future directions. *Int. J. Precis. Eng. Manuf. Technol.*, 3, 111–128.
- Jessop-Fabre, M.M. and Sonnenschein, N. (2019) Improving reproducibility in synthetic biology. *Front. Bioeng. Biotechnol.*, 7, 18.
- Sandoval, C.M., Ayson, M., Moss, N., Lieu, B., Jackson, P., Gaucher, S.P., Horning, T., Dahl, R.H., Denery, J.R., Abbott, D.A. et al. (2014) Use of pantothenate as a metabolic switch increases the genetic stability of farnesene producing *Saccharomyces cerevisiae*. *Metab. Eng.*, 25, 215–226.
- Ye, H., Baba, M.D.-E., Peng, R.-W. and Fussenegger, M. (2011) A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science* (80-), 332, 1565–1568.
- Hillson, N., Caddick, M., Cai, Y., Carrasco, J.A., Chang, M.W., Curach, N.C., Bell, D.J., Le Feuvre, R., Friedman, D.C., Fu, X. et al. (2019) Building a global alliance of biofoundries. *Nat. Commun.*, 10, 2040.
- Wehrs, M., de Beaumont-Felt, A., Goranov, A., Harrigan, P., de Kok, S., Lieder, S., Vallandingham, J. and Tyner, K. (2020) You get what you screen for: on the value of fermentation characterization in high-throughput strain improvements in industrial settings. *J. Ind. Microbiol. Biotechnol.*, doi: 10.1007/s10295-020-02295-3.
- Carbonell, P., Radivojevic, T. and García Martín, H. (2019) Opportunities at the intersection of synthetic biology, machine learning, and automation. *ACS Synth. Biol.*, 8, 1474–1477.
- Angermueller, C., Pärnamaa, T., Parts, L. and Stegle, O. (2016) Deep learning for computational biology. *Mol. Syst. Biol.*, 12, 878.
- Wu, Z., Kan, S.B.J., Lewis, R.D., Wittmann, B.J. and Arnold, F.H. (2019) Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. USA*, 116, 8852–8858. doi: 10.1073/pnas.1901979116.
- Yang, K.K., Wu, Z. and Arnold, F.H. (2019) Machine-learning-guided directed evolution for protein engineering. *Nat. Methods*, 16, 687–694. doi: 10.1038/s41592-019-0496-6.
- Amidi, A., Amidi, S., Vlachakis, D., Megalooikonomou, V., Paragios, N. and Zacharaki, E.I. (2018) EnzyNet: enzyme classification using 3D convolutional neural networks on spatial representation. *PeerJ*, 6, e4750.
- Kim, G.B., Kim, W.J., Kim, H.U. and Lee, S.Y. (2020) Machine learning applications in systems metabolic engineering. *Curr. Opin. Biotechnol.*, 64, 1–9.
- Costello, Z. and Martin, H.G. (2018) A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst. Biol. Appl.*, 4, 19.
- Opgenorth, P., Costello, Z., Okada, T., Goyal, G., Chen, Y., Gin, J., Benites, V., De Raad, M., Northen, T.R., Deng, K. et al. (2019) Lessons from two design-build-test-learn cycles of dodecanol production in *Escherichia coli* aided by machine learning. *ACS Synth. Biol.*, 8, 1337–1351.
- Karim, A.S., Dudley, Q.M., Juminaga, A., Yuan, Y., Crowe, S.A., Heggestad, J.T., Garg, S., Abdalla, T., Grubbe, W.S., Rasor, B.J. et al. (2020) In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat. Chem. Biol.*, 16, 912–919. doi: 10.1038/s41589-020-0559-0.
- Halper, S.M., Hossain, A. and Salis, H.M. (2020) Synthesis success calculator: predicting the rapid synthesis of DNA fragments with machine learning. *ACS Synth. Biol.*, 9, 1563–1571.
- Reis, A.C., Halper, S.M., Vezeau, G.E., Cetnar, D.P., Hossain, A., Clauer, P.R. and Salis, H.M. (2019) Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat. Biotechnol.*, 37, 1294–1301.
- Jervis, A.J., Carbonell, P., Vinaixa, M., Dunstan, M.S., Hollywood, K.A., Robinson, C.J., Rattray, N.J.W., Yan, C., Swainston, N., Currin, A. et al. (2019) Machine learning of designed translational control allows predictive pathway optimization in *Escherichia coli*. *ACS Synth. Biol.*, 8, 127–136.
- Hiscock, T.W. (2019) Adapting machine-learning algorithms to design gene circuits. *BMC Bioinformatics*, 20, 214.

26. Clomburg, J.M., Crumbley, A.M. and Gonzalez, R. (2017) Industrial biomanufacturing: the future of chemical production. *Science* (80-), 355, aag0804.
27. Sun, C., Theodoropoulos, C. and Scrutton, N.S. (2020) Techno-economic assessment of microbial limonene production. *Bioresour. Technol.*, 300, 122666.
28. Carbonell, P., Gök, A., Shapira, P. and Faulon, J.-L. (2016) Mapping the patent landscape of synthetic biology for fine chemical production pathways. *Microb. Biotechnol.*, 9, 687–695.
29. Campodonico, M.A., Sukumara, S., Feist, A.M. and Herrgård, M.J. (2018) Computational methods to assess the production potential of bio-based chemicals. In: Jensen, M.K., Keasling, J.D. (eds). *Methods in Molecular Biology* (Clifton, N.J.), Vol. 1671. *Synthetic Pathways*, Springer, pp. 97–116.
30. Yoshikuni, Y., Dietrich, J.A., Nowroozi, F.F., Babbitt, P.C. and Keasling, J.D. (2008) Redesigning enzymes based on adaptive evolution for optimal function in synthetic metabolic pathways. *Chem. Biol.*, 15, 607–618.
31. Lin, G.-M., Warden-Rothman, R. and Voigt, C.A. (2019) Retrosynthetic design of metabolic pathways to chemicals not found in nature. *Curr. Opin. Syst. Biol.*, 14, 82–107.
32. Wang, L., Dash, S., Ng, C.Y. and Maranas, C.D. (2017) A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol.*, 2, 243–252.
33. Hadadi, N. and Hatzimanikatis, V. (2015) Design of computational retrosynthesis tools for the design of de novo synthetic pathways. *Curr. Opin. Chem. Biol.*, 28, 99–104.
34. Campodonico, M.A., Andrews, B.A., Asenjo, J.A., Palsson, B.O. and Feist, A.M. (2014) Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. *Metab. Eng.*, 25, 140–158.
35. Delépine, B., Duigou, T., Carbonell, P. and Faulon, J.-L. (2018) RetroPath2.0: a retrosynthesis workflow for metabolic engineers. *Metab. Eng.*, 45, 158–170.
36. Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S. and Kanehisa, M. (2010) PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res.*, 38, W138–W143.
37. Cho, A., Yun, H., Park, J., Lee, S. and Park, S. (2010) Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Syst. Biol.*, 4, 35.
38. Segler, M.H.S., Preuss, M. and Waller, M.P. (2018) Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, 555, 604–610.
39. Duigou, T., Du Lac, M., Carbonell, P. and Faulon, J.-L. (2019) RetroRules: a database of reaction rules for engineering biology. *Nucleic Acids Res.*, 47, D1229–D1235.
40. Rahman, S.A., Cuesta, S.M., Furnham, N., Holliday, G.L. and Thornton, J.M. (2014) EC-BLAST: a tool to automatically search and compare enzyme reactions. *Nat. Methods*, 11, 171–174.
41. Carbonell, P., Wong, J., Swainston, N., Takano, E., Turner, N.J., Scrutton, N.S., Kell, D.B., Breitling, R. and Faulon, J.-L. (2018) Selenzyme: enzyme selection tool for pathway design. *Bioinformatics*, 34, 2153–2154.
42. Arnold, F.H. (2018) Enzymes by evolution: bringing new chemistry to life. *Mol. Front. J.*, 02, 9–18.
43. Webb, S. (2018) Deep learning for biology. *Nature*, 554, 555–557.
44. Li, Y., Wang, S., Umarov, R., Xie, B., Fan, M., Li, L. and Gao, X. (2018) DEEPre: sequence-based enzyme EC number prediction by deep learning. *Bioinformatics*, 34, 760–769.
45. Ryu, J.Y., Kim, H.U. and Lee, S.Y. (2019) Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. *Proc. Natl. Acad. Sci. USA*, 116, 13996–14001. doi: 10.1073/PNAS.1821905116.
46. Mazurenko, S., Prokop, Z. and Damborsky, J. (2020) Machine learning in enzyme engineering. *ACS Catal.*, 10, 1210–1223.
47. Segall-Shapiro, T.H., Sontag, E.D. and Voigt, C.A. (2018) Engineered promoters enable constant gene expression at any copy number in bacteria. *Nat. Biotechnol.*, 36, 352–358.
48. Jervis, A., Carbonell, P., Taylor, S., Sung, R., Dunstan, M.S., Robinson, C.J., Breitling, R., Takano, E. and Scrutton, N.S. (2019) SelProm: a queryable and predictive expression vector selection tool for *Escherichia coli*. *ACS Synth. Biol.*, 8, 1478–1483.
49. Jones, J.A., Vernacchio, V.R., Lachance, D.M., Lebovich, M., Fu, L., Shirke, A.N., Schultz, V.L., Cress, B., Linhardt, R.J. and Koffas, M.A.G. (2015) ePathOptimize: a combinatorial approach for transcriptional balancing of metabolic pathways. *Sci. Rep.*, 5, 1–10.
50. Salis, H.M. (2011) The ribosome binding site calculator. In: C., Voigt (ed). *Methods in Enzymology*, Vol. 498; 'Synthetic Biology, Part B Computer aided design and DNA assembly'. Elsevier, pp. 19–42.
51. Na, D. and Lee, D. (2010) RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics*, 26, 2633–2634.
52. Jeschek, M., Gerngross, D. and Panke, S. (2016) Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.*, 7, 1163.
53. Swainston, N., Dunstan, M., Jervis, A.J., Robinson, C.J., Carbonell, P., Williams, A.R., Faulon, J.-L., Scrutton, N.S., Kell, D.B. and Kelso, J. (2018) PartsGenie: an integrated tool for optimizing and sharing synthetic biology parts. *Bioinformatics*, 34, 2327–2329.
54. Roehner, N., Young, E., Voigt, C.A., Gordon, D.B. and Densmore, D. (2016) Double Dutch: a tool for designing combinatorial libraries of biological systems. *ACS Synth. Biol.*, 5, 507–517. doi: 10.1021/acssynbio.5b00232.
55. Hossain, A., Lopez, E., Halper, S.M., Cetnar, D.P., Reis, A.C., Strickland, D., Klavins, E. and Salis, H.M. (2020) Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. *Nat. Biotechnol.*, doi: 10.1038/s41587-020-0584-2.
56. Casini, A., Chang, F.-Y., Eluere, R., King, A.M., Young, E.M., Dudley, Q.M., Karim, A., Pratt, K., Bristol, C., Forget, A. et al. (2018) A pressure test to make 10 molecules in 90 days: external evaluation of methods to engineer biology. *J. Am. Chem. Soc.*, 140, 4302–4316.
57. Robinson, C.J., Carbonell, P., Jervis, A.J., Yan, C., Hollywood, K.A., Dunstan, M.S., Currin, A., Swainston, N., Spiess, R., Taylor, S. et al. (2020) Rapid prototyping of microbial production strains for the biomanufacture of potential materials monomers. *Metab. Eng.*, 60, 168–182.
58. Dunstan, M.S., Robinson, C.J., Jervis, A.J., Yan, C., Carbonell, P., Hollywood, K.A., Currin, A., Swainston, N., Le Feuvre, R., Micklefield, J. et al. (2020) Engineering *Escherichia coli* towards de novo production of gatekeeper (2S)-flavanones: naringenin, pinocembrin, eriodictyol and homoeriodictyol. *Synth. Biol.*, doi: 10.1093/synbio/ysaa012.
59. Tianqi, L., Simmons, T.L., Bohnsack, D.A., Mackay, M.E., Smith, M.R. and Baker, G.L. (2007) Synthesis of polymandelide: a degradable polylactide derivative with polystyrene-like properties. *Macromolecules*, 40, 6040–6047.
60. Martinková, L. and Křen, V. (2018) Biocatalytic production of mandelic acid and analogues: a review and comparison with chemical processes. *Appl. Microbiol. Biotechnol.*, 102, 3893–3900.

- 
61. Chemat, F., Vian, M.A., Ravi, H.K., Khadhraoui, B., Hilali, S., Perino, S. and Tixier, A.S.F. (2019) Review of alternative solvents for green extraction of food and natural products: panorama, principles, applications and prospects. *Molecules*, 24, 3007.
62. Amer, M., Wojcik, E.Z., Sun, C., Hoeven, R., Hoeven, R., Hughes, J.M.X., Faulkner, M., Yunus, I.S., Tait, S., Johannissen, L.O. et al. (2020) Low carbon strategies for sustainable bio-alkane gas production and renewable energy. *Energy Environ. Sci.*, 13, 1818–1831.
63. Zheng, P., Li, C., Wang, N., Li, J. and An, Q. (2019) The potential of pervaporation for biofuel recovery from fermentation: an energy consumption point of view. *Chin. J. Chem. Eng.*, 27, 1296–1306.
64. Patraşcu, I., Bîldea, C.S. and Kiss, A.A. (2018) Eco-efficient downstream processing of biobutanol by enhanced process intensification and integration. *ACS Sustain. Chem. Eng.*, 6, 5452–5461.
65. Köpke, M. and Simpson, S.D. (2020) Pollution to products: recycling of ‘above ground’ carbon by gas fermentation. *Curr. Opin. Biotechnol.*, 65, 180–189.
66. Takors, R., Kopf, M., Mampel, J., Bluemke, W., Blombach, B., Eikmanns, B., Bengelsdorf, F.R., Weuster-Botz, D. and Dürre, P. (2018) Using gas mixtures of CO, CO and H<sub>2</sub> as microbial substrates: the do’s and don’ts of successful technology transfer from laboratory to production scale. *Microb. Biotechnol.*, 11, 606–625.
67. Blank, L.M., Narancic, T., Mampel, J., Tiso, T. and O’Connor, K. (2020) Biotechnological upcycling of plastic waste and other non-conventional feedstocks in a circular economy. *Curr. Opin. Biotechnol.*, 62, 212–219.
68. Papaioannou, E.H., Liakopoulou-Kyriakides, M. and Karabelas, A.J. (2016) Natural origin lycopene and its “Green” downstream processing. *Crit. Rev. Food Sci. Nutr.*, 56, 686–709.