Next-generation intelligent laboratories for materials design and manufacturing

Xiting Peng and Xiaonan Wang*

The contradiction between the importance of materials to modern society and their slow development process has led to the development of multiple methods to accelerate materials discovery. The recently emerged concept of intelligent laboratories integrates the developments in fields of high-throughput experimentation, automation, theoretical computing, and artificial intelligence to form a system that can autonomously carry out designed experiments and make scientific discoveries. We present the basic concepts and the foundations of this new research paradigm, demonstrate its typical application scenarios through case studies, and envision a collaborative human-machine meta laboratory in the future.

Introduction

From clean energy to human welfare and even national security, advanced materials are critical to the development of society. However, the fact is that the materials development process is very long. A new material needs to go through many steps from discovery to deployment, often taking decades.[1] For example, Li-ion batteries have been studied on a large scale since the early 1960s and were not commercialized until the 1990s.[2] On the other hand, the complexity of the materials is increasing with the continuous exploration of improved materials.[3] The combination of just a few parameters can lead to a huge search space, which is impossible to be fully explored by traditional manual experiments.

A variety of technologies have been applied to accelerate materials development, which encompasses the four paradigms of scientific research (**Figure 1**). High-throughput experimental techniques have led to continuous improvements in the scale and efficiency of experiments. Theoretical calculation methods with databases storing the calculation results allow screening to be performed rapidly on computers.[4] Data science-related technologies, especially machine learning, have led to data-driven discoveries in materials science.[5] In this context, the concept of a self-driving laboratory or autonomous experimentation system has been proposed. It is an integration of the advances previously mentioned, and has already shown strong potential in fields such as drug discovery,[6–8] solar-energy materials,[9–11] organic synthesis,[7,12] etc.

Here, we reinterpret these efforts into a more general concept of "intelligent laboratories." In using this term, we are not focusing on a specific experimental method or facility, but on a system that combines all aspects of materials research to provide a systematic solution to the materials development process (Figure 1). True intelligence should be born out of the collaboration of humans and robots. A future intelligent laboratory involves not only the replacement of human hands with robotic arms, but, more importantly, the enhancement of the human brain with efficient algorithms and computing power. The experiments are designed and executed in a closed-loop, autonomous manner, with the human effort focusing on initializing the system and analyzing the final results. This separation of labor allows researchers to devote more effort to the aspects that really require innovation and insight.

There are a variety of apparent benefits of an intelligent laboratory. It allows for uninterrupted, high-throughput parallel experiments that are far more efficient than those performed by humans. It also minimizes operational faults and errors,

Xiting Peng, Department of Chemical Engineering, Tsinghua University, Beijing, China; pxt22@mails.tsinghua.edu.cn
Xiaonan Wang, Department of Chemical Engineering, Tsinghua University, Beijing, China; Key Laboratory of Industrial Biocatalysis (Tsinghua University), Ministry of Education, Beijing, China; wangxiaonan@tsinghua.edu.cn
*Corresponding author
doi:10.1557/s43577-023-00481-z

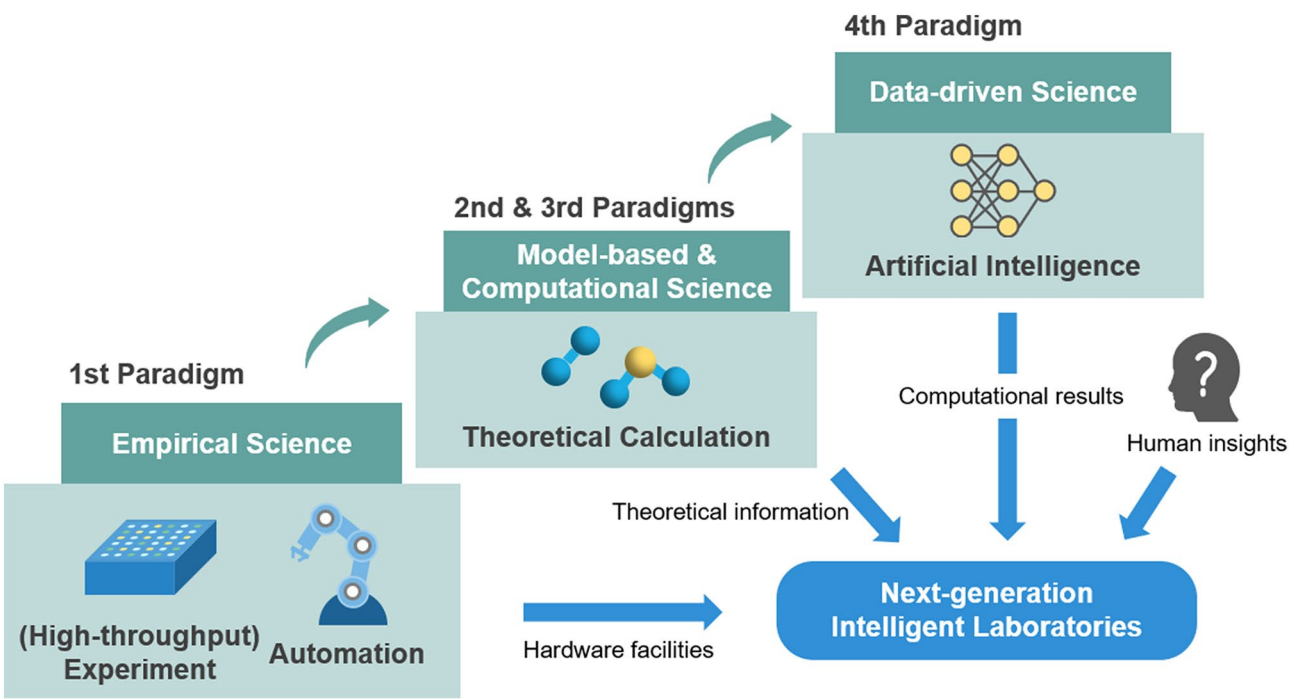

Figure 1. Demonstration of the four research paradigms linking to next-generation intelligent laboratories. The intelligent laboratory is not a new technology, but a system in which all aspects of materials research are involved.

which can generate standardized, high-quality data. Considering a robot can use different vision and motion systems than a human, it is also more flexible for handling particular substances such as photosensitive materials and hazardous chemicals.[11] In addition, researchers can be physically separated from the experimental equipment and conduct experiments in a remote, shared manner, making "cloud chemistry"[13] or even "meta laboratories" possible, which is particularly relevant in the post-pandemic era.

This article aims to present the concept of future intelligent laboratories and to explain what an intelligent laboratory is and why it is important. We will demonstrate the current application scenarios, advantages, and limitations of intelligent laboratories with several case studies, with an outlook on how they can further advance materials design and manufacturing.

Fundamentals of an intelligent laboratory

In 2009, a robot scientist named Adam was proposed to discover functional genes in *Saccharomyces cerevisiae*.[14] It uses automated hardware combined with a software system to iteratively formulate hypotheses, design experiments, and interpret results without human intervention. A mobile robotic chemist developed in 2020 takes a more advanced approach, using Bayesian optimization to guide a dexterous roaming robot through the laboratory to autonomously find improved photocatalysts for hydrogen production from water.[11] Both cases follow the workflow described in **Figure 2**. First, the researcher searches for the area of interest and builds the required hardware and software environment. Then, for the task to be studied, the researcher defines the parameters and objectives to be optimized and selects a proper optimization framework. The system is also provided with initial data or other *a priori* knowledge. After these initializations, the entire system begins to operate automatically. A model is used to infer the results of each test, and a planning algorithm incorporates the model to determine how the next experiment should be performed and then drives the automated device to perform the test. The results obtained are passed back to the computer to update the model and determine the next experiment. After hours or days, the iterations are completed, and all the experimental data are aggregated into a conclusion. In another case, the initialization step can also be done by an AI model automatically reading and generalizing the existing human knowledge.[15]

The emergence of this research paradigm was not happening overnight, but was gradually transformed. Specifically, an intelligent laboratory often integrates the following components that will be discussed next:

1. Automated, modular hardware facilities. It may be a set of modified commercial instruments, an *in situ* characterization system, and/or a robotic arm;
2. A computational core (brain) of the intelligent laboratory for analyzing experimental results and planning the next experiments;
3. A software system to serve as an interface between the researcher, the computational tools, and the hardware facilities; and

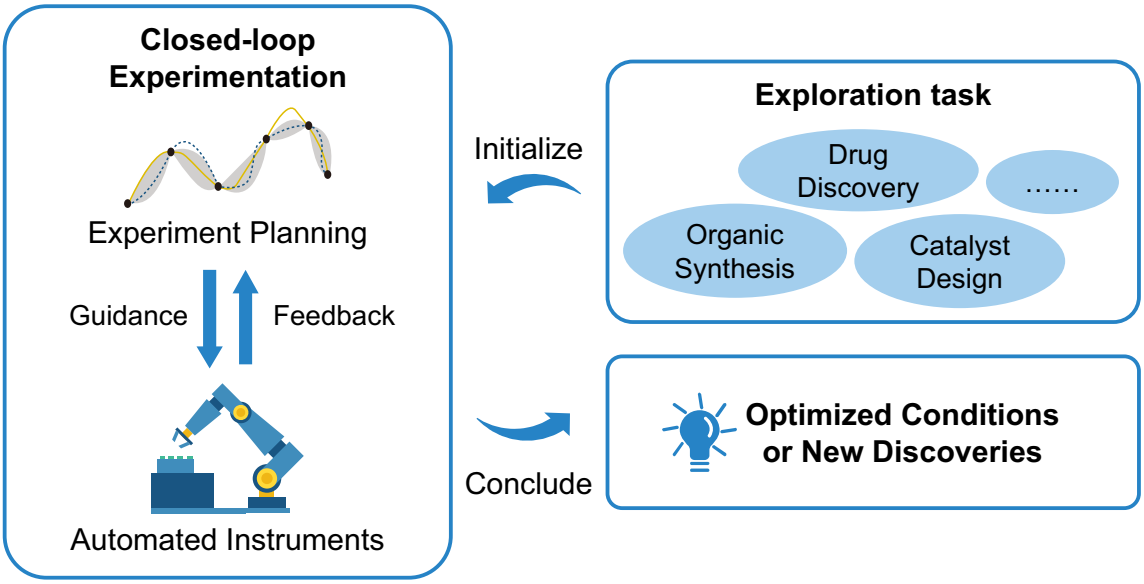

Figure 2. The general workflow of an intelligent laboratory. The definition of the system is done manually, whereas the process of experiment planning and conducting is fully autonomous.

4. Other strategies to assist experiments, such as introducing the results of theoretical calculations to make predictions physics-informed.

High-throughput experiment with experiment planning

Advanced experimental methods are the cornerstone of research, and intelligent laboratories are dedicated to how to do experiments smarter. Early research for materials design was often Edison-like. Researchers relied on personal experience and intuition to discover new materials by trial and error, which was often a time-consuming and unpredictable process. Starting in the 1990s, the combinatorial method emerged in the field of materials science,[16] drawing on the idea of large-scale synthesis and characterization in biological research.[17] The combined experiments mean to increase the speed of materials synthesis and testing through miniaturized and parallel synthesis methods and high-throughput characterization methods. A library containing a large number of materials with their properties is created, based on which one could select the best material or explore the relationship between structure and properties. This advancement has substantially increased the speed of discovery and deployment of new materials. However, if we rely only on blind and random testing with a large number of experiments, much of the experimental information obtained will be redundant and is still insignificant in the face of the vast possibility space. To improve the value of each test, some planning strategy should be used to explicitly choose the next experiment to be performed.

Statistical design of experiment (DOE) is an early method applied to experimental planning, which analyzes the interrelationships between multiple independent variables and their effects on the dependent variable by systematic sampling and testing. It is a powerful method when the number of independent variables is small and the relationships between them are relatively simple.[18] Subsequently, early artificial intelligence (AI) methods and advanced optimization algorithms began to be applied to materials science for more complex problems. For example, the global single-objective optimization algorithm SNOBFIT[19] can perform optimization without a theoretical model and has been used in the optimization of chemical reactions.[12] Some multi-objective optimization methods, such as TSEMO[20] and MOAL,[21] have also been developed for the tradeoff between conflicting objectives. With the development of artificial intelligence, active learning[22,23] and reinforcement learning[24] techniques have started to be used in materials research. Bayesian optimization is a popular black-box optimization method in recent times, which uses a surrogate function to approximate an unknown experimental response surface and an acquisition function to select the parameter for the next measurement. The selection of parameters takes into account the most promising regions and the regions with high uncertainty (i.e., the optimal solution is found while gradually improving the accuracy of the surrogate function). Several studies[23,25,26] have demonstrated the effectiveness of these advanced planning methods. By using them, the number of experiments and time consumed were reduced significantly (by several to tens of times) compared to the traditional grid-based approach or random search.

Theoretical calculation

The theoretical calculation can be considered as a further expansion of experimental throughput, and was highlighted in some national-level materials acceleration initiatives such as

the Materials Genome Initiative (MGI).[1] With high-throughput calculations,[27] researchers can explore a larger range of chemical spaces without conducting experiments and perform virtual screening directly on the computer.[28] Unlike the predictions given by data-driven machine learning models, the results of theoretical calculations are derived directly from fundamental physical principles and are considered to be more reliable and physically explainable, but also take longer to compute. Therefore, large computational materials databases[29–31] have been created to store the results of theoretical calculations and reduce repetitive calculations. Machine learning methods are used by introducing surrogate models trained on the fly to speed up the search process carried out by theoretical calculations.[32–34] In our previous work, theoretical calculations can also be added directly to the active learning loops to enable self-improving materials discovery.[35]

Also, the theoretical calculation can be combined with other techniques to develop new research methods, such as using the results of theoretical calculations as pretraining data for machine learning prediction models,[29] or combing it with high-throughput screening to take advantage of both of their advantages.[36]

AI for intelligent laboratories

Artificial intelligence techniques can assist materials science in many ways, including property prediction, molecular design, theoretical calculations acceleration, and more, which have been fully discussed in previous reviews.[5,37,38] In addition, some advanced AI techniques can also be applied to materials discovery. Active learning and data augmentation can be used to cope with data scarcity and imbalance, and transfer learning enables fast and resource-saving learning on similar tasks, as we demonstrated respectively in the design of strain sensors and metal–organic framework-based membranes.[39,40] The biggest contribution of AI for intelligent laboratories lies in AI planning, which exists as the "computational brain" of the system to guide the next experiment (as described in the section "High-throughput experiment with experiment planning").

Considering that the deep learning (DL) method, another important part of AI, often involves building and training data-intensive models that do not seem to match the goal of "more efficient experimentation," it is difficult to directly apply it in intelligent laboratories.[41] But it can still play a role in various aspects. For example, we used machine vision to analyze scanning probe microscopy images, which can improve the speed and accuracy of chiral molecule detection and classification, allowing more characterization tools to be added to automated closed-loop experiments.[42] Another example is machine reading for information sorting from literature. It allows integration of existing knowledge and suggests more promising hypotheses for experiments.[15,43]

Automated experimental hardware

Automation technology is the bridge between physical instruments and the computational core, and it determines the scope of tasks to which we can apply the closed-loop autonomous research paradigm. It can be stressed that automation and digitization make a lot of sense: standardized operation and comprehensive data recording. Materials are complex and small changes in synthesis conditions or some factors that may be overlooked during manual experimentation (e.g., laboratory temperature and humidity, sample purity) may have a large impact on the experiment. On the other hand, although only the most successful experimental results are reported in publications, failed experiments often contain chemical intuition, and the collection and use of this information can also aid in materials discovery.[44,45] Therefore, the future laboratory needs to digitally record and integrate all information from experiments.

Although the robot operating system (ROS)[46] has reached a mature level in the early years and automated platforms have been broadly applied in areas such as drug discovery, its full application to other fields is still limited by diverse methodology and costs. Sokol et al.[46] evaluated the difficulty of automating commonly used materials synthesis and characterization techniques for solar fuel production. More easily parallelized materials synthesis processes and more in-line characterization instruments still deserve to be developed.[47] On the other hand, software packages such as ChemOs,[48] ESCALATE,[49] etc., have been developed to facilitate the deployment and operation of an intelligent laboratory for a wider range of people.

Application scenarios of an intelligent laboratory

Exploration of materials formulations and synthesis conditions

A classical application scenario for intelligent laboratories can be concluded as achieving the best property by optimizing the formulations or synthesis conditions of materials, whether they are carbon nanotubes,[50] gold nanoparticles,[51,52] quantum dots,[53] thin-film materials,[9] or metastable materials.[22] The synthesis-characterization cycle of materials is often fixed and can be done on a single robotic platform or even a single instrument. The difficulty lies in the large number of processing parameters available with insufficient knowledge about their relationships, which can be explored more efficiently using an automated experimentation platform driven by AI.

In 2016, Nikolaev et al.[50] reported the first Autonomous Research System (ARES) for controlling the growth rate of carbon nanotubes (CNTs). CNTs were synthesized using a chemical vapor deposition process on small pillars of silicon and were characterized by *in situ* Raman spectroscopy for the growth rate. Synthesis conditions, including temperature, pressure, and gas composition, were optimized by the suggestion from a random forest model. It was observed that the predicted and experiment growth rate got closer as the number of completed experiments increased, and a target growth rate was achieved after 600 experiments without human intervention.

A platform such as this not only allows for faster completion of specific experimental tasks, but its independence from *a priori* knowledge also allows for full exploration of previously unexplored design space, which may lead to more generalized results. For example, in the design of battery electrolytes, it has been shown that mixing salts may have a positive impact on performance, but the huge design space brings difficulty to the experiment, and there is still no guarantee of fidelity of calculations from first principles. Dave et al.[54] then applied the research framework of autonomous experiments, using pumping devices and electrochemical tools to form a simple and inexpensive platform for testing electrochemical performance. After 140 experiments with three- and four-salt aqueous electrolytes, an electrolyte that was contrary to the usual design principles was obtained. In another example, because crystal structure characterization using x-ray diffraction (XRD) is very expensive, Ament et al.[22] used AI agents and an active learning strategy to establish a synthesis phase diagram of a Bi-O system with as few experiments as possible, thus discovering the conditions for stabilizing δ-Bi_2O_3 at room temperature.

Intelligent laboratories for organic synthesis

The research paradigm of intelligent laboratories can also be applied to organic synthesis. For each reaction, a large number of repetitive experiments are required to find the optimal reaction conditions in order to obtain the maximum yield and minimize the byproducts. Angello et al.[55] devised a closed-loop optimization process for general reaction conditions of the heteroaryl Suzuki-Miyaura coupling reaction. Substrates were first selected by data mining techniques, and then a combination of uncertainty-minimizing ML and robotic experimentation was used to perform iterative tests to optimize the reaction conditions. Reaction conditions that doubled the average yield of the currently widely used conditions were found, and the generality also showed statistically significant increases compared to the previously reported conditions. This workflow improves the efficiency of reaction optimization, whereas organic synthesis processes often involve different types of reactions, and the optimization of different reactions requires reconfigurability of the system. A plug-and-play, continuous-flow chemical synthesis system was developed by Bédard et al.[12] This system is highly integrated, allowing the selection and configuration of different modular unit operations and analytical methods as needed in a simple way. Optimization of reaction conditions was performed automatically and a user-friendly interface was provided to reach high usability. The system's capabilities have been demonstrated in several widely used single- and two-step reactions.

Another important task is the synthesis of organic compounds on demand. Coley et al.[7] first integrated advances in the computer-aided synthesis planning (CASP) and automated synthesis platforms to construct an automatic flow synthesis platform. Information from millions of published synthetic routes was extracted by data mining and machine learning methods to generate recommended synthetic routes. A self-configurable system using a robotic arm for on-demand assembly of unit operations and reagent lines is used to perform the experiments. Fifteen small molecules that vary greatly in structure and synthetic complexity were synthesized automatically. It should be noted that the training of models in this system is still a one-time process that relies on historical data, whereas some specific reaction conditions need to be defined manually due to the lack of data. These limitations can hopefully be solved by the introduction of the previously mentioned closed-loop optimization process, which will enable a step further toward the automation of all aspects of organic synthesis.

Fully autonomous laboratory for catalyst design

Although building an autonomous experimentation platform for a specific task is not that challenging, the range of applications is limited. Several efforts aim to build a fully autonomous laboratory that can perform multiple research tasks. Building such a platform is certainly time-consuming, but once such a laboratory is fully operational, the benefits can still be significant.

Burger et al.[11] presented a landmark work named mobile robotic chemist. A mobile robotic platform with a robotic arm was used to mimic human behavior. With the help of laser scanning combined with touch feedback for fine positioning, the platform can move freely in the laboratory like a human, grasp and transfer sample vials, and operate commercial instruments with no physical hardware modifications. The complete experimental process was performed by eight workstations. By using a batch Bayesian algorithm to optimize the catalyst formulations, a photocatalyst mixture with six times higher activity than the initial formulations was identified after 688 experiments.

In 2022, Zhu et al.[15] developed AI-Chemist, an AI chemistry laboratory that implements the whole process of scientific research autonomously. For an exploration task, a large number of papers were first read and summarized by AI, and scientific hypotheses were generated. A computational module for performing theoretical calculations was used to provide physics-based data for the prediction model. A Bayesian algorithm was used to guide the search space. Experiments were carried out by a mobile robot that shuttled between 14 modified workstations, including synthesis, characterization, and performance tests. The effectiveness of the platform was demonstrated in the design of a non-noble metal oxygen evolution reaction electrocatalyst. AI-Chemist explored in a search space with 553,401 options and found a composition ratio that was far from the best samples of trial-and-error experiments but performed better.

Future directions of intelligent laboratories

Scientific research is always about the data obtained from the experiments, not the physical samples themselves. Thus researchers can be physically separated from their

experimental instruments and samples. The COVID-19 pandemic has impacted scientific research globally, and the need to digitalize laboratory operations in an era of social distance has become more urgent. The concept of intelligent laboratories offers the prospect that we can redefine scientific research using digital technologies such as cloud computing, the Internet of Things, and blockchain, building "meta laboratories" that fully map physical laboratories. Future experimental research can be conducted in large, well-resourced "virtual institutes" accessible to anyone without leaving home. The sharing of experiment resources and data will expand the use of experimental instruments and chemicals, making science more democratized, as in the astronomy and high-energy physics communities.[41]

As a preliminary stage, some commercial cloud laboratories have been built, such as Strateos (https://strateos.com/), dedicated to drug discovery and synthetic biology, and Emerald Cloud Lab (https://www.emeraldcloudlab.com/), which can perform various *in vitro* biology or chemistry experiments. Researchers can control multiple devices through a software interface or web browser without purchasing their own instruments. It makes biological and chemical experiments more like computational experiments, with each step of the experimental operation being translated into explicit and detailed code and carried out in an invariable environment, which provides for reproducibility. Unlike contract research organizations (CROs), which provide research outsourcing in the biomedical field, how experiments are performed in this research mode is still user-defined. Although the founder of the Emerald Cloud Lab admits that access to its equipment is not cheap, he believes that it can be a money-saving option compared to the cost of purchasing the equipment.[56] However, a research model in which robots are merely used to replace some operation processes is not the ultimate goal. Humans are good at causal analysis and reasoning, whereas robots are good at cognition and execution. In a future laboratory, humans and robots should collaborate to get the most out of each other.[57] Such a promising research paradigm would be applied more often in future materials development to provide solutions to current thorny problems.

Acknowledgments

This work has been supported by the Scientific and Technological Innovation 2030 - "New Generation Artificial Intelligence" Major Project (2022AAA010062) by the Ministry of Science and Technology in China.

Data availability

Not applicable.

Code availability

Not applicable.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, *APL Mater.* **1**(1), 011002 (2013). https://doi.org/10.1063/1.4812323
2. M. Winter, B. Barnett, K. Xu, *Chem. Rev.* **118**(23), 11433 (2018). https://doi.org/10.1021/acs.chemrev.8b00422
3. Z. Rao, P.-Y. Tung, R. Xie, Y. Wei, H. Zhang, A. Ferrari, T.P.C. Klaver, F. Körmann, P.T. Sukumar, A. Kwiatkowski da Silva, Y. Chen, Z. Li, D. Ponge, J. Neugebauer, O. Gutfleisch, S. Bauer, D. Raabe, *Science* **376**(6615), 78 (2022). https://doi.org/10.1126/science.abo4940
4. S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, *Comput. Mater. Sci.* **58**, 218 (2012). https://doi.org/10.1016/j.commatsci.2012.02.005
5. K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **559**, 547 (2018). https://doi.org/10.1038/s41586-018-0337-2
6. G. Schneider, *Nat. Rev. Drug Discov.* **17**, 97 (2018). https://doi.org/10.1038/nrd.2017.232
7. C.W. Coley, D.A. Thomas III, J.A.M. Lummiss, J.N. Jaworski, C.P. Breen, V. Schultz, T. Hart, J.S. Fishman, L. Rogers, H. Gao, R.W. Hicklin, P.P. Plehiers, J. Byington, J.S. Piotti, W.H. Green, A.J. Hart, T.F. Jamison, K.F. Jensen, *Science* **365**(6453), eaax1566 (2019). https://doi.org/10.1126/science.aax1566
8. B. Desai, K. Dixon, E. Farrant, Q. Feng, K.R. Gibson, W.P. van Hoorn, J. Mills, T. Morgan, D.M. Parry, M.K. Ramjee, C.N. Selway, G.J. Tarver, G. Whitlock, A.G. Wright, *J. Med. Chem.* **56**(7), 3033 (2013). https://doi.org/10.1021/jm400099d
9. B.P. MacLeod, F.G.L. Parlane, T.D. Morrissey, F. Häse, L.M. Roch, K.E. Dettelbach, R. Moreira, L.P.E. Yunker, M.B. Rooney, J.R. Deeth, V. Lai, G.J. Ng, H. Situ, R.H. Zhang, M.S. Elliott, T.H. Haley, D.J. Dvorak, A. Aspuru-Guzik, J.E. Hein, C.P. Berlinguette, *Sci. Adv.* **6**(20), aaz8867 (2020). https://doi.org/10.1126/sciadv.aaz8867
10. B.P. MacLeod, F.G.L. Parlane, C.C. Rupnow, K.E. Dettelbach, M.S. Elliott, T.D. Morrissey, T.H. Haley, O. Proskurin, M.B. Rooney, N. Taherimakhsousi, D.J. Dvorak, H.N. Chiu, C.E.B. Waizenegger, K. Ocean, M. Mokhtari, C.P. Berlinguette, *Nat. Commun.* **13**, 995 (2022). https://doi.org/10.1038/s41467-022-28580-6
11. B. Burger, P.M. Maffettone, V.V. Gusev, C.M. Aitchison, Y. Bai, X. Wang, X. Li, B.M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R.S. Sprick, A.I. Cooper, *Nature* **583**, 237 (2020). https://doi.org/10.1038/s41586-020-2442-2
12. A.-C. Bédard, A. Adamo, K.C. Aroh, M.G. Russell, A.A. Bedermann, J. Torosian, B. Yue, K.F. Jensen, T.F. Jamison, *Science* **361**(6408), 1220 (2018). https://doi.org/10.1126/science.aat0650
13. R.A. Skilton, R.A. Bourne, Z. Amara, R. Horvath, J. Jin, M.J. Scully, E. Streng, S.L.Y. Tang, P.A. Summers, J. Wang, E. Pérez, N. Asfaw, G.L.P. Aydos, J. Dupont, G. Comak, M.W. George, M. Poliakoff, *Nat. Chem.* **7**, 1 (2015). https://doi.org/10.1038/nchem.2143
14. R.D. King, J. Rowland, S.G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L.N. Soldatova, A. Sparkes, K.E. Whelan, A. Clare, *Science* **324**(5923), 85 (2009). https://doi.org/10.1126/science.1165620
15. Q. Zhu, F. Zhang, Y. Huang, H. Xiao, L. Zhao, X. Zhang, T. Song, X. Tang, X. Li, G. He, B. Chong, J. Zhou, Y. Zhang, B. Zhang, J. Cao, M. Luo, S. Wang, G. Ye, W. Zhang, X. Chen, S. Cong, D. Zhou, H. Li, J. Li, G. Zou, W. Shang, J. Jiang, Y. Luo, *Natl. Sci. Rev.* **9**(10), nwac190 (2022). https://doi.org/10.1093/nsr/nwac190
16. X.-D. Xiang, X. Sun, G. Briceño, Y. Lou, K.-A. Wang, H. Chang, W.G. Wallace-Freedman, S.-W. Chen, P.G. Schultz, *Science* **268**(5218), 1738 (1995). https://doi.org/10.1126/science.268.5218.1738
17. S.J. Pollack, J.W. Jacobs, P.G. Schultz, *Science* 234(4783), 1570 (1986). https://doi.org/10.1126/science.3787262
18. L. Harmon, *J. Mater. Sci.* **38**, 4479 (2003). https://doi.org/10.1023/a:1027325400459
19. W. Huyer, A. Neumaier, *ACM Trans. Math. Softw.* **35**(2), 9 (2008). https://doi.org/10.1145/1377612.1377613
20. E. Bradford, A.M. Schweidtmann, A. Lapkin, *J. Glob. Optim.* **71**, 407 (2018). https://doi.org/10.1007/s10898-018-0609-2
21. N. Peremezhney, E. Hines, A. Lapkin, C. Connaughton, *Eng. Optim.* **46**(11), 1593 (2014). https://doi.org/10.1080/0305215X.2014.881997
22. S. Ament, M. Amsler, D.R. Sutherland, M.-C. Chang, D. Guevarra, A.B. Connolly, J.M. Gregoire, M.O. Thompson, C.P. Gomes, R.B. van Dover, *Sci. Adv.* **7**(51), abg4930 (2021). https://doi.org/10.1126/sciadv.abg4930

23. A.G. Kusne, H. Yu, C. Wu, H. Zhang, J. Hattrick-Simpers, B. DeCost, S. Sarker, C. Oses, C. Toher, S. Curtarolo, A.V. Davydov, R. Agarwal, L.A. Bendersky, M. Li, A. Mehta, I. Takeuchi, *Nat. Commun.* **11**, 5966 (2020). https://doi.org/10.1038/s41467-020-19597-w
24. J. Li, J. Li, R. Liu, Y. Tu, Y. Li, J. Cheng, T. He, X. Zhu, *Nat. Commun.* **11**, 2046 (2020). https://doi.org/10.1038/s41467-020-15728-5
25. A. Dave, J. Mitchell, S. Burke, H. Lin, J. Whitacre, V. Viswanathan, *Nat. Commun.* **13**, 5454 (2022). https://doi.org/10.1038/s41467-022-32938-1
26. A.E. Gongora, B. Xu, W. Perry, C. Okoye, P. Riley, K.G. Reyes, E.F. Morgan, K.A. Brown, *Sci. Adv.* **6**(15), aaz1708 (2020). https://doi.org/10.1126/sciadv.aaz1708
27. S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, *Nat. Mater.* **12**, 191 (2013). https://doi.org/10.1038/nmat3568
28. E.O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, A. Aspuru-Guzik, *Annu. Rev. Mater. Sci.* **45**, 195 (2015). https://doi.org/10.1146/annurev-matsci-070214-020823
29. J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, *JOM* **65**, 1501 (2013). https://doi.org/10.1007/s11837-013-0755-4
30. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, *APL Mater.* **1**(1), 011002 (2013). https://doi.org/10.1063/1.4812323
31. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, *Comput. Mater. Sci.* **58**, 227 (2012). https://doi.org/10.1016/j.commatsci.2012.02.002
32. T.L. Jacobsen, M.S. Jørgensen, B. Hammer, *Phys. Rev. Lett.* **120**, 026102 (2018). https://doi.org/10.1103/PhysRevLett.120.026102
33. Z.W. Ulissi, A.J. Medford, T. Bligaard, J.K. Nørskov, *Nat. Commun.* **8**, 14621 (2017). https://doi.org/10.1038/ncomms14621
34. A.A. Peterson, *J. Chem. Phys.* **145**(7), 074106 (2016). https://doi.org/10.1063/1.4960708
35. S. Xu, J. Li, P. Cai, X. Liu, B. Liu, X. Wang, *J. Am. Chem. Soc.* **143**(47), 19769 (2021). https://doi.org/10.1021/jacs.1c08211
36. R.S. Ferreira, A. Simeonov, A. Jadhav, O. Eidam, B.T. Mott, M.J. Keiser, J.H. McKerrow, D.J. Maloney, J.J. Irwin, B.K. Shoichet, *J. Med. Chem.* **53**(13), 4891 (2010). https://doi.org/10.1021/jm100488w
37. Z. Yao, Y. Lum, A. Johnston, L.M. Mejia-Mendoza, X. Zhou, Y. Wen, A. Aspuru-Guzik, E.H. Sargent, Z.W. Seh, *Nat. Rev. Mater.* (2022). https://doi.org/10.1038/s41578-022-00490-5
38. J. Li, K. Lim, H. Yang, Z. Ren, S. Raghavan, P.-Y. Chen, T. Buonassisi, X. Wang, *Matter* **3**(2),393 (2020). https://doi.org/10.1016/j.matt.2020.06.011
39. H. Yang, J. Li, K.Z. Lim, C. Pan, T. Van Truong, Q. Wang, K. Li, S. Li, X. Xiao, M. Ding, T. Chen, X. Liu, Q. Xie, P.V.Y. Alvarado, X. Wang, P.-Y. Chen, *Nat. Mach. Intell.* **4**, 84 (2022). https://doi.org/10.1038/s42256-021-00434-8
40. J. Guan, T. Huang, W. Liu, F. Feng, S. Japip, J. Li, J. Wu, X. Wang, S. Zhang, *Cell Rep. Phys. Sci.* **3**(5), 100864 (2022). https://doi.org/10.1016/j.xcrp.2022.100864
41. E. Stach, B. DeCost, A.G. Kusne, J. Hattrick-Simpers, K.A. Brown, K.G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C.P. Gomes, J.M. Gregoire, A. Mehta, J. Montoya, E. Olivetti, C. Park, E. Rotenberg, S.K. Saikin, S. Smullin, V. Stanev, B. Maruyama, *Matter* **4**(9), 2702 (2021). https://doi.org/10.1016/j.matt.2021.06.036
42. J. Li, M. Telychko, J. Yin, Y. Zhu, G. Li, S. Song, H. Yang, J. Li, J. Wu, J. Lu, X. Wang, *J. Am. Chem. Soc.* **143**(27), 10177 (2021). https://doi.org/10.1021/jacs.1c03091
43. S.H.M. Mehr, M. Craven, A.I. Leonov, G. Keenan, L. Cronin, *Science* **370**(6512), 101 (2020). https://doi.org/10.1126/science.abc2986
44. P. Raccuglia, K.C. Elbert, P.D.F. Adler, C. Falk, M.B. Wenny, A. Mollo, M. Zeller, S.A. Friedler, J. Schrier, A.J. Norquist, *Nature* **533**, 73 (2016). https://doi.org/10.1038/nature17439
45. S.M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K.C. Stylianou, B. Smit, *Nat. Commun.* **10**, 539 (2019). https://doi.org/10.1038/s41467-019-08483-9
46. K.P. Sokol, V. Andrei, *Nat. Rev. Mater.* **7**, 251 (2022). https://doi.org/10.1038/s41578-022-00432-1
47. H. Xu, J. Lin, Q. Liu, Y. Chen, J. Zhang, Y. Yang, M.C. Young, Y. Xu, D. Zhang, F. Mo, *Chem* **8**(12), 3202 (2022). https://doi.org/10.1016/j.chempr.2022.08.008
48. L.M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L.P.E. Yunker, J.E. Hein, A. Aspuru-Guzik, *PLoS One* **15**(4), e0229862 (2020). https://doi.org/10.1371/journal.pone.0229862
49. I.M. Pendleton, G. Cattabriga, Z. Li, M.A. Najeeb, S.A. Friedler, A.J. Norquist, E.M. Chan, J. Schrier, *MRS Commun.* **9**, 846 (2019). https://doi.org/10.1557/mrc.2019.72
50. P. Nikolaev, D. Hooper, F. Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto, B. Maruyama, *NPJ Comput. Mater.* **2**, 16031 (2016). https://doi.org/10.1038/npjcompumats.2016.31
51. Y. Jiang, D. Salley, A. Sharma, G. Keenan, M. Mullin, L. Cronin, *Sci. Adv.* **8**(40), abo2626 (2022). https://doi.org/10.1126/sciadv.abo2626
52. D. Salley, G. Keenan, J. Grizou, A. Sharma, S. Martín, L. Cronin, *Nat. Commun.* **11**, 2771 (2020). https://doi.org/10.1038/s41467-020-16501-4
53. R.W. Epps, M.S. Bowen, A.A. Volk, K. Abdel-Latif, S. Han, K.G. Reyes, A. Amassian, M. Abolhasani, *Adv. Mater.* **32**(30), 2001626 (2020). https://doi.org/10.1002/adma.202001626
54. A. Dave, J. Mitchell, K. Kandasamy, H. Wang, S. Burke, B. Paria, B. Póczos, J. Whitacre, V. Viswanathan, *Cell Rep. Phys. Sci.* **1**(12), 100264 (2020). https://doi.org/10.1016/j.xcrp.2020.100264
55. N.H. Angello, V. Rathore, W. Beker, A. Wołos, E.R. Jira, R. Roszak, T.C. Wu, C.M. Schroeder, A. Aspuru-Guzik, B.A. Grzybowski, M.D. Burke, *Science* **378**(6618), 399 (2022). https://doi.org/10.1126/science.adc8743
56. C. Arnold, *Nature* **606**, 612 (2022). https://doi.org/10.1038/d41586-022-01618-x
57. C.A. Nicolaou, C. Humblet, H. Hu, E.M. Martin, F.C. Dorsey, T.M. Castle, K.I. Burton, H. Hu, J. Hendle, M.J. Hickey, J. Duerksen, J. Wang, J.A. Erickson, *ACS Med. Chem. Lett.* **10**(3), 278 (2019). https://doi.org/10.1021/acsmedchemlett.8b00488 □

Publisher's note

Xiting Peng is a doctoral candidate in the Department of Chemical Engineering at Tsinghua University, China. She received her bachelor's degree in bioengineering from Tianjin University, China, in 2022. Her research focuses on artificial intelligence methods for materials laboratories. Peng can be reached by email at pxt22@mails.tsinghua.edu.cn.

Xiaonan Wang is an associate professor in the Department of Chemical Engineering at Tsinghua University, China. She received her BEng degree from Tsinghua University in 2011, and PhD degree from the University of California, Davis, in 2015. After working as a postdoctoral research associate at Imperial College London, Wang joined the National University of Singapore (NUS) as an assistant professor in 2017, and later became an adjunct associate professor. She is leading a Smart Systems Engineering research group at NUS and Tsinghua and is also the deputy director of the Accelerated Materials Development Programme. Her awards include the 2023 ACS Sustainable Chemistry & Engineering Lectureship Award, World's Top 2% Scientists and the American Institute of Chemical Engineers-Singapore Local Section (AIChE-SLS) Outstanding Young Principal Investigator. Wang can be reached by email at wangxiaonan@tsinghua.edu.cn.