PERSPECTIVES

Accelerating the discovery of materials for clean energy in the era of smart automation

Daniel P. Tabor ⓘ, Loïc M. Roch ⓘ, Semion K. Saikin ⓘ, Christoph Kreisbeck, Dennis Sheberla ⓘ, Joseph H. Montoya ⓘ, Shyam Dwaraknath ⓘ, Muratahan Aykol ⓘ, Carlos Ortiz, Hermann Tribukait, Carlos Amador-Bedolla ⓘ, Christoph J. Brabec ⓘ, Benji Maruyama, Kristin A. Persson ⓘ and Alán Aspuru-Guzik ⓘ

Abstract | The discovery and development of novel materials in the field of energy are essential to accelerate the transition to a low-carbon economy. Bringing recent technological innovations in automation, robotics and computer science together with current approaches in chemistry, materials synthesis and characterization will act as a catalyst for revolutionizing traditional research and development in both industry and academia. This Perspective provides a vision for an integrated artificial intelligence approach towards autonomous materials discovery, which, in our opinion, will emerge within the next 5 to 10 years. The approach we discuss requires the integration of the following tools, which have already seen substantial development to date: high-throughput virtual screening, automated synthesis planning, automated laboratories and machine learning algorithms. In addition to reducing the time to deployment of new materials by an order of magnitude, this integrated approach is expected to lower the cost associated with the initial discovery. Thus, the price of the final products (for example, solar panels, batteries and electric vehicles) will also decrease. This in turn will enable industries and governments to meet more ambitious targets in terms of reducing greenhouse gas emissions at a faster pace.

Advanced materials affect most, if not all, aspects of life today. They are crucial for technologies ranging from energy generation, transmission and storage to water filtration, power electronics, transportation and aerospace[1–6]. These areas all require materials that satisfy increasingly demanding performance specifications. Innovation is imperative to reach these goals and can be stimulated by the development of novel integrated artificial intelligence (AI) algorithms and robotics into fully autonomous platforms.

The timelines for materials discovery, development and deployment are long, and the process is capital intensive. Typically, new materials technologies reach the market after 10–20 years of basic and applied research[7]. Platforms that integrate AI with automated and robotized synthesis and characterization have the potential to accelerate the entire materials discovery and innovation process to reduce this time by an order of magnitude.

The transformation of the current materials discovery pipelines into the proposed integrated platforms requires commitment from key players, ranging from governments and academic research institutions to large industries and capital providers[8]. Knowledge transfer across different specialized industries represents a challenge to bring in industry and private sector players. However, this is not only a challenge but also an opportunity. As the discovery processes for advanced materials accelerate, the potential economic benefits will grow exponentially. Thus, private sector stakeholders that join these efforts early will presumably have a first-mover advantage; they will have the know-how to adjust and obtain a larger share of these growing benefits[9].

Building a multidisciplinary workforce for the discovery, production and integration of advanced materials requires efforts and leadership from academia, governments and industry[10–12]. The continuous support of research and development initiatives, such as the Materials Genome Initiative[13,14], will aid in the development and deployment of a discovery workforce that is ready for the challenges ahead. International coalitions around particular topics could advance the agenda and produce results more effectively.

An example of such an international collaboration is Mission Innovation, a coalition of 22 countries plus the European Union that have committed to doubling their investments in clean energy innovation by 2021. Mission Innovation focuses on seven Innovation Challenges: smart grids, off-grid access to electricity, carbon capture use and sequestration, sustainable biofuels, converting sunlight into fuels, clean energy materials, and affordable heating and cooling of buildings. The focus in this Perspective is on the efforts relating to the Clean Energy Materials Innovation Challenge; however, accelerating the discovery of high-performance novel materials is important for all seven of the Innovation Challenges. In line with the Paris Agreement, adopted in December 2015, the aim is to limit the increase in the global average temperature to less than 2 °C (REF.[15]). In this context, clean energy innovations and disruptive technological breakthroughs are essential to meet the reduction targets for greenhouse gas emissions and even more ambitious targets in the near future[16]. The goal of the Clean Energy Materials Innovation Challenge is to propel materials discovery and to develop new high-performance, low-cost clean energy solutions.

In this Perspective, we provide our vision for the next generation of integrated AI approaches towards autonomous materials discovery. Bridging the gaps between independent technologies that are essential to materials discovery and incorporating them into a single platform will alleviate the clash that often happens between theory and experiment.

First, we briefly discuss advances in AI and then provide an overview of the main applications of materials for the clean energy sector. Next, we explore state-of-the-art automated procedures for materials discovery, with a focus on machine learning. The field of organic materials is the farthest along in many of the areas required for an integrated platform, but along the way, we point out some of the notable advances in both inorganic materials and nanomaterials. Finally, we conclude and provide our vision for the next generation of integrated AI approaches towards autonomous materials discovery, which will emerge within the next 5–10 years.

Advances in AI

Scientific discoveries are usually associated with an insight: the act of intuitively seeing a phenomenon, which contrasts with systematic mechanistic learning. This is despite the fact that most of our discoveries are based on extensive preliminary studies. Insight is considered to be an exclusively human attribute, whereas systematic exploration is connected with automated platforms. However, this gap between a creative and intuitive targeted search and a systematic exploration continues to narrow as automated platforms become increasingly sophisticated and are able to process more complex information[17,18]. In addition to systematic screenings of large databases and building chemical structures according to a set of preprogramed rules, today's platforms can update the rules for analysing the available information and even search for more information that helps them to make specific decisions[19]. In this case, it is natural to expect that in the very near future such platforms will be able to not only predict the properties of materials but also test hypotheses by designing structures and characterizing them, becoming autonomous.

It has already been demonstrated that combinatorial optimization procedures provide faster screening of the molecular space than traditional approaches based on intuition[20]. The pharmaceutical and chemical industries, as well as academic research environments, use these methods for the design of new molecules, reactions and materials[21–26]. However, combinatorial chemical synthesis makes an exhaustive search of the multidimensional molecular space out of reach[26]. As such, the community needs a more rational approach for exploration of this large space; this is where machine learning comes into play.

The recent progress in machine learning and statistical inference methods can be viewed as a revolution in AI. Most machine learning methods, such as neural networks and Bayesian optimization, were developed decades ago but have not found widespread use until recently[27,28]. Basic research in AI continues to be backed by governments, industry, and public and private research institutes[29]. Today, machine learning methods are behind many commercial applications, such as Internet searches, natural language translation, and image and speech recognition.

Recently, an upgraded version of AlphaGo, the Go-playing program from Google, which in part uses deep neural networks (DNNs) and reinforcement learning as key algorithms, beat the top human Go player[30]. Moreover, its playing style inspired other Go players. Since mid-2017, Cisco has used a multilayer supervised algorithm based on machine learning to analyse encrypted traffic (that is, HTTPS). This algorithm helps identify malware communication through passive monitoring and yields enhanced incident responses[31]. In economics, machine learning has also started to emerge, notably to predict economic growth[32], to quantify predictive performance[33] and to anticipate customer behaviour[34]. While recent progress is making its mark in non-scientific endeavours, the application of machine learning in science and medicine is also emerging: assisting physicians in interpreting computer-based medical images, processing biomedical signals and learning from patient data[35–37]. In late 2017, a DNN was successfully shown to enable the reconstruction of perceptual and subjective images from the activity of human brains[38].

Within the context of this Perspective, the application of several machine learning methods to computational chemistry has recently bourgeoned[39–41]. From the representation of aromaticity and conjugation in general[42] to the prediction of protein–ligand affinities[43], there is increasing interest in using DNNs and convolutional neural networks in a wide range of applications. Hybrid learning models, which combine different approaches to leverage their respective strengths, have shown great promise. Examples of hybrid learning models include Bayesian deep learning[44], Bayesian conditional generative adversarial networks[45] and deep Bayesian optimization[46]. The latter has been successfully applied to reverse engineer chemical reactions to quantitatively

and qualitatively reproduce observed behaviours[46]. Machine-learning-based algorithms have also been intensively used to bypass expensive static[47] and dynamic[48–50] ab initio electronic structure calculations. Although exploration and discovery are more challenging than interpolation or optimization for AI, recent algorithmic developments show substantial promise for making advances in these areas. To overcome these challenges of inverse design in computational chemistry and to explore the open-ended chemical space, autoencoders and generative adversarial networks have emerged as powerful tools to generate novel molecular structures with desirable properties tailored to specific needs[51–55]. This progress is only the beginning of the integrated materials discovery revolution.

The key component of an autonomous discovery approach lies in the synergy between machine learning and robotics. One might ask, "Why are robots better?" A cursory analysis of humans versus robots shows some clear advantages for the latter. First, robots can operate in more adverse conditions. This can be seen even outside chemistry; for example, robots were sent to Mars decades before humans. For chemistry, this ability can translate to procedures that are subject to high temperatures and/or pressures, toxic solvents and highly exothermic processes. Robots also excel at providing unbiased and reproducible routes towards materials discovery. For example, it was recently demonstrated that not only do machine-learning-based algorithms cover an application space of polyoxometalates approximately six times larger than a human approach but they also increase the accuracy of prediction by a relative value of 6.9%[56]. Recently, an autonomous infrastructure for the optimization of chemical reaction conditions through the Deep Reaction Optimizer (DRO), an algorithm based on deep reinforcement learning, was reported[57]. Furthermore, robots are better at recording reaction procedures independently of their outcome and reduce waste by rigorously following the stoichiometry of the experiments[18]. Robots also provide a natural platform for scaling chemical experiments, reducing the cost per experiment. In 2009, a hypothetico-deductive 'robot scientist', named Adam, was developed that could autonomously perform experiments, devise hypotheses and design experiments to validate the hypotheses in the area of functional genomics[58].

Clean energy generation and storage

To accelerate the transition to a low-carbon economy, the deployment of clean energy technologies, along with possibilities for carbon utilization, storage and capture, must be implemented[59–61]. On the one hand, several commercial entities[62–65] are actively developing technologies to capture and possibly convert CO_2 into fuels or feedstock chemicals at emissions sources. On the other hand, valorization of CO_2 to chemical energy carriers is key to the realization of a low-carbon economy[66–68]. However, valorization is unlikely to outpace CO_2 emissions from our current global demand for energy in the near term, and thus the aforementioned multifaceted approach, which includes the development of advanced materials for energy conversion and storage, is a necessity.

Below, we outline five types of clean energy technology — catalysis, photovoltaics (PVs), thermoelectrics, energy-efficient materials and energy storage solutions (FIG. 1) — and the relevance of automated materials discovery in these areas.

Catalysis

The chemical industry is among the largest consumers of energy, accounting for roughly 15% of the total US energy consumption[69]. Alternative technologies that produce chemicals and fuels from renewable feedstocks could greatly mitigate emissions and may enable continued use of our existing energy consumption infrastructure. The key materials science need in this domain is in the development of catalysts and chemical processes that can convert earth-abundant molecules into fuels and chemicals with costs comparable to those of fossil fuels. The difficulty of breaking and reforming chemical bonds in these relatively inert feedstocks, such as CO_2, H_2O and N_2, makes achieving this goal a sizeable challenge.

Recent efforts have demonstrated that routes towards the valorization of CO_2 exist[70–75]. In May 2017, the world's first commercial plant, built by Climeworks, for capturing CO_2 directly from air opened in Zürich, Switzerland[76] (FIG. 1a). Valorization processes are highly dependent on catalysis and often use expensive or scarce metals, such as Pt, as catalysts, which prohibits notable scale up. In response to these shortcomings, high-throughput materials screening[77,78] that couples theory and experiment has been used to discover earth-abundant materials, including MoS_2 and NiGa alloys for electrocatalytic H_2 evolution and for CH_3OH synthesis from CO_2, respectively[79,80]. Ultimately, these limitations on both cost and catalytic efficiency require considerable advances in our ability to conceive and synthesize nanostructured materials[81].

To address these challenges, machine learning and automation are increasingly being used to refine the theoretical approach to materials discovery. Bayesian optimization, for example, has emerged as a powerful tool for the rapid characterization of catalytic surfaces[82–84]. Genetic algorithms have also shown promise in accelerating compositional searches for catalysts for the synthesis of CH_3OH (REF.[85]) and structural optimization of catalytic clusters on support materials[86]. Modelling complex reaction mechanisms on surfaces may be greatly facilitated using techniques from machine learning to recognize the most relevant reaction pathways to sample for a given material[83]. In addition, machine-learned interatomic potentials, such as those in the smooth overlap of atomic positions (SOAP) framework[87], may prove useful for investigating reaction dynamics, nanoparticle formation, surface equilibria and other catalytic phenomena on larger length scales for which quantum chemical methods are currently intractable[88].

One major obstacle to realizing a more fully integrated and autonomous system of catalyst discovery is the lack of a robust and community-wide framework of informatics with respect to data related to surface characterization and reactivity. Simulation and experimental data on catalytic activity, adsorption energy and surface energy are inherently more complex than the data for many other models of computational materials science (including elements such as faceting, reconstructions, molecular degrees of freedom and solvation). As a result, the community-wide approach is often ad hoc and tailored too narrowly to specific chemistries to be broadly generalized. High-throughput approaches are beginning to be developed[84], but the computational cost of collecting larger data sets often necessitates strict structural or mechanistic assumptions about either the investigated materials or the catalytic cycle[89,90]. A more unified approach that enables researchers to contribute data to a community-wide database might make efforts to use machine learning and automation more viable. Such efforts have begun, for example, with the community databases CatApp[91] and crystalium[92], but substantial improvements to data infrastructure in catalysis are likely necessary to fully realize the potential of AI to accelerate the discovery of useful catalysts[93].

Photovoltaics

Photovoltaics (PVs) have experienced tremendous growth in electricity generation in the past decade owing to a dramatic reduction in costs over the past 5 years[94], leading to their integration into many parts of society, including building façades (FIG. 1b). Today, PVs are the least expensive technology to use as the energy source

Heat source

n   p

Heat sink

Visible light

Near-infrared light
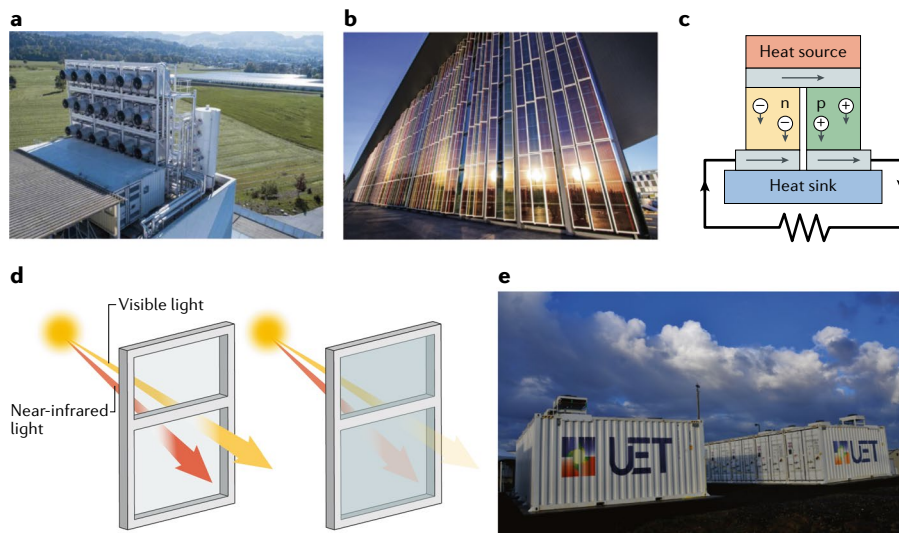
Fig. 1 | **Examples of clean energy generation and storage technologies. a** | The world's first commercial plant for CO_2 capture, Zürich, Switzerland[76]. **b** | First demonstration of the integration of dye-sensitized solar cells into a building façade, SwissTech Convention Center, Switzerland[288]. **c** | Scheme of a thermoelectric device[289]. **d** | Schematic representation of electrochromic windows[290]. **e** | Vanadium flow (V-flow) batteries produced by UniEnergy Technologies in South Africa[291]. Panel **a** is reproduced with permission from Climeworks. Panel **b** is reproduced with permission from Solaronix. Panel **c** is adapted from C.M. Cullen, CC-BY-2.5. Panel **e** is used with the permission of UniEnergy Technologies, LLC.

for newly built power stations in many countries. However, the scale up of current PV technology is fairly slow and expensive and requires enormous financial investments into manufacturing facilities. Thus, PVs remain far from providing a sufficient fraction of today's total energy demand. There is therefore much scope to develop materials for a PV technology with improved scalability to guarantee a sustainable and continuous growth rate of the worldwide PV production[95]. One of the main challenges in developing novel PV materials is the time to market. The uptake of PV technology is driven by efficiency, lifetime and costs, and the complexity of fulfilling all three parameters in parallel considerably delays the market implementation of technologies. On average, it takes 25–35 years from the first report before a novel material is manufactured in relevant volume for commercial application.

New research initiatives need to focus on technology breakthroughs rather than on small incremental power conversion improvements. Radically new concepts for novel materials and processes need to address several important aspects of PV panels. These include increasing power conversion efficiencies beyond the Shockley–Queisser limit, suppressing materials degradation under harsh conditions and improving recyclability of the cells. The environmental effects and costs of manufacturing, installation and maintenance also need to be considered. All these parameters are heavily dependent on the physical properties of the active and supportive materials, as well as the interfaces between the materials in a PV device, making the optimization of existing materials and the development of novel materials a highly complex and challenging mission. There are a number of exciting new materials concepts that may succeed the existing Si technology that currently dominates the PV market[94]. Solution-processed technologies, such as organometal halide perovskites[96–99] and organic PVs (OPVs)[100–105], are highly attractive owing to their reasonably high performances (in terms of power conversion efficiencies) of 22% and 14%, respectively, as well as their uniquely low-cost production by low-temperature printing and coating methods. Further enhancing the performance and especially the environmental stability of OPVs requires the invention of novel material classes and innovative concepts based on additives that provide enhanced stability against oxidation and photo-oxidation[106]. The targeted design and synthesis of stable interface layers require the development of new types of doping procedures, with corresponding new types of dopants, that are less expensive than the vacuum-based procedures currently used. Solution-processed barrier materials with high crystallinity and low oxygen and water-vapour transmission rates need to be invented, as thin-film vacuum packaging might not be compatible with the future cost structure of PVs. The successful development of these aspects would position OPVs as a highly competitive technology for markets that are tailored to their unique strengths, such as building-integrated PVs or non-grid-connected PVs[107–109].

New materials for PVs could be investigated more quickly using a combination of machine-learning-driven electronic structure calculations. Recently, such a method was used to predict Flory–Huggins parameters for PV materials[110]. Furthermore, the combinatorial screening of polymer:fullerene blends as inks for PVs was demonstrated on a customized inkjet platform[111]. Also encouraging was a recent report on the automated high-throughput synthesis and characterization of novel organometal halide perovskites for PVs[112]. The transfer of a rapid precipitation-based synthesis process to a high-throughput platform enabled microcrystalline semiconductors to be synthesized within seconds. By scanning several hundreds of compositions, the optimal stoichiometry with the best efficiency and stability was identified. We believe that these findings lay the foundations for the design of the next generation of composites that combine high efficiency with excellent stability.

Thermoelectrics

Vast amounts of excess heat are generated daily, worldwide. According to a Lawrence Livermore National Laboratory energy-flow analysis, more than 66% of generated energy is lost, much of it as waste heat[113]. One solution is to harness the current levels of industrial waste heat for use in thermoelectric devices (FIG. 1c), that is, to convert heat into electricity. Although there has been an increase in the performance of thermoelectric materials in the past decade, continued improvements are still needed for devices to be viable for waste heat recovery[114]. Finding materials with high thermoelectric performance, as measured by the thermoelectric figure of merit (ZT), is challenging because of the interdependence of the electrical conductivity, the Seebeck coefficient (or thermoelectric sensitivity) and the electronic thermal conductivity.

The highest performance bulk inorganic materials have a $ZT \approx 2$, which corresponds to 15–20% heat into electricity conversion[115]. These high-performance inorganic materials often contain rare-earth elements, making the devices cost prohibitive for wide-scale deployment. However, the advent of robust first-principles methods and high-performance computing resources has opened up the way for in silico materials design. For example, novel inorganic thermoelectric materials have recently been designed using a combination of high-throughput first-principles calculations and data-mined substitution rules[116–120]. There are also ongoing efforts to identify alternative inorganics[116,120] or materials composed of more abundant organics. Currently, the maximum ZT for organic materials has reached only about one-quarter of the ZT of the best inorganic materials[121]. However, organic[122], metal–organic[123] and hybrid materials offer different pathways to optimize ZT (especially around room temperature) and may lead to device geometries that were not previously possible with inorganic counterparts. Accelerated discovery in this field is necessary for the development of new inexpensive materials with a high ZT that are scalable in terms of both materials availability and manufacturing. This is imperative in order for thermoelectric devices to become commercially viable and suitable for large-scale deployment.

Energy-efficient materials

Reducing energy consumption in the residential and commercial building sectors is a major factor in the transition to a low-carbon economy. In 2010, these sectors contributed to about 30% of all energy-related CO_2 emissions[124]. Within these sectors, there are many opportunities for novel materials that enable key innovations, such as smart windows (FIG. 1d) or sophisticated cooling strategies based on phase-change materials[125–127], to contribute to a reduction in energy consumption. For example, phase-change materials can be incorporated into a concrete matrix (so-called Thermocrete)[127] that stores excess heat during the day and releases it during cool nights. By contrast, super-insulating materials, such as tailored aerogels[128–130] or nanoscale engineered foams[131], could pave the way to energy-efficient, passive houses. Smart windows[132–135] allow for active control of the transmittance of infrared radiation while remaining transparent. These windows are estimated to yield a reduction in the

energy demand of up to 40% compared with traditional static windows[136].

Electrochromic materials have a long history, and various material classes, such as viologens (both in solution or as polymer films)[137–140] and conjugated conducting polymers[141–143], have been extensively studied. For a more comprehensive overview, we refer the interested reader to REFS[144,145] and the references therein. Recently, metal–organic frameworks (MOFs) have emerged as promising candidates for electrochromic materials[146–150]. For example, when MOF-74 was functionalized with redox-active naphthalene diimide ligands, the material exhibited fast switching from transparent to dark[148]. The major advantage of MOFs lies in the tunability of their material properties, for example, by adding linkers or through post-synthetic modifications. This flexibility makes MOFs extremely versatile but calls for efficient tools to scan the accessible chemical space of MOFs efficiently. Virtual computational screening, genetic algorithms and machine learning models have started to be successfully applied to predict and optimize properties of MOFs[151–155]. For example, a data set comprising 130,398 MOFs[154] was used to train various machine learning models, such as random forest and support vector machines. These models yielded accurate predictions of CH_4 adsorption capacities with errors reported to be as low as 7.18% for cross-validation[155]. Although previous works have focused mainly on the virtual screening of MOFs for gas storage, we note that the developed tools can, in principle, be transferred to other contexts, such as electrochromic devices.

There are a few commercial electrochromic products available, but broader adoption requires additional research efforts to address challenges such as durability, performance and switching time, which in many ways are parallel to the problems outlined above for PVs.

Energy storage solutions
Many renewable energy generation technologies are intermittent in nature, and thus the development of robust, reliable and efficient energy storage technologies is central to large-scale deployment and market penetration. This area of research has attracted intense attention and has spawned over a dozen new technologies, with each catering to a different application based on requirements such as power and total storage capacity[1]. The state of

research in energy storage technologies can be roughly divided into two different categories: the first contains technologies that present mostly engineering challenges at this point (for example, compressed air or pumped hydro) and the second is composed of technologies that could be revolutionized by a fundamental breakthrough in materials development. This latter category includes, for example, metal batteries[156], flow batteries[157] and supercapacitors[158]. The Li-ion battery, with its ubiquitous presence in modern society, is probably the best-known energy storage technology. Research in post-Li-ion batteries has been active over the past decades[156,159,160], but many technologies with positive results in laboratory tests have not yet been commercialized. The same layered electrode materials that were discovered decades ago for use in Li-ion batteries are still the workhorses of the industry; most of the improvements have been achieved by engineering the cell and fine-tuning the materials chemistry. Nevertheless, the ability to rapidly calculate and data-mine structure–function correlations has had an impact on the research of novel energy storage materials. For example, new classes of phosphate and carbonophosphate Li-ion cathodes were generated as 'virtual' compounds by substituting ions in known compounds from the Inorganic Crystal Structure Database, using a data-mined model[161], to form new possible compounds containing Li, P, O and a redox-active metal. The properties (for example, voltage, capacity, stability and safety) of the resulting candidate compounds were then evaluated by first-principles calculations, and as a result, new materials were synthesized and characterized on the basis of these calculations[162,163]. Machine learning and data-mining-assisted methods, such as those used for multi-battery systems optimization, state-of-charge monitoring, remaining-useful-life prognostics and smart load balancing, also offer great promise for maximizing the utilization and cycle life of Li-ion cells[164–167].

Redox-flow batteries, despite exhibiting relatively low energy density, offer a promising solution for grid-based energy storage because their power output and energy capacity can be engineered independently of each other and because they use electrolytes that, in principle, can last much longer than those used in Li-ion batteries. Vanadium flow batteries (FIG. 1e) have greatly improved since their invention in the 1970s but are fundamentally limited by the rarity of vanadium[168]. Recent research

in the flow-battery community has involved trying to replace the standard vanadium pair of electrolytes with less-expensive electrolytes, but no pair of molecules has been discovered to date that has all the necessary properties[168–173]. Supercapacitors have promise as potential storage solutions for high-power, short-duration applications[158]. Recent developments in supercapacitors have focused on a variety of materials for both the electrode (ranging from graphene and carbon nanotubes to metal oxides) and the electrolyte material.

All these storage solutions could experience further materials development in the near future. Given the molecular nature of many of these materials, the machine-learning-driven structure–property relationship methods outlined above could be applied to drive research forward even in the short term.

The closed-loop approach
Although an integrated, or closed-loop, approach for materials discovery has yet to be demonstrated, there has been sufficient progress in designing the individual, mandatory components. The workflow of a closed-loop approach (FIG. 2) begins with identifying an appropriate application space within all of chemical space for the problem at hand; this corresponds to library generation. Next, the application space is narrowed to a set of promising leads, using various levels of theory. There is a first feedback loop at this stage, and on the basis of the identified promising leads, the application space is adapted for the next step. From the promising leads, the reaction space, which is defined as the subspace of synthesizable molecules, is generated by the automated synthetic planning of reaction pathways. If there are no known methods found, then the rules for generating the library are updated through another feedback loop to overcome the existing constraints. The reaction space is then narrowed to the robotics space by evaluating the synthesizability of candidate molecules subject to the hardware constraints of the robotics solution. This robotics space is then subject to testing using automated synthesis and characterization, and the performance here feeds back directly into both the current robotics space and library generation, which is the beginning of the loop. The feedback loops between the outlined steps are the key characteristics of the approach that differentiate AI materials discovery machinery from the standard high-throughput processes already widely adopted by industry.
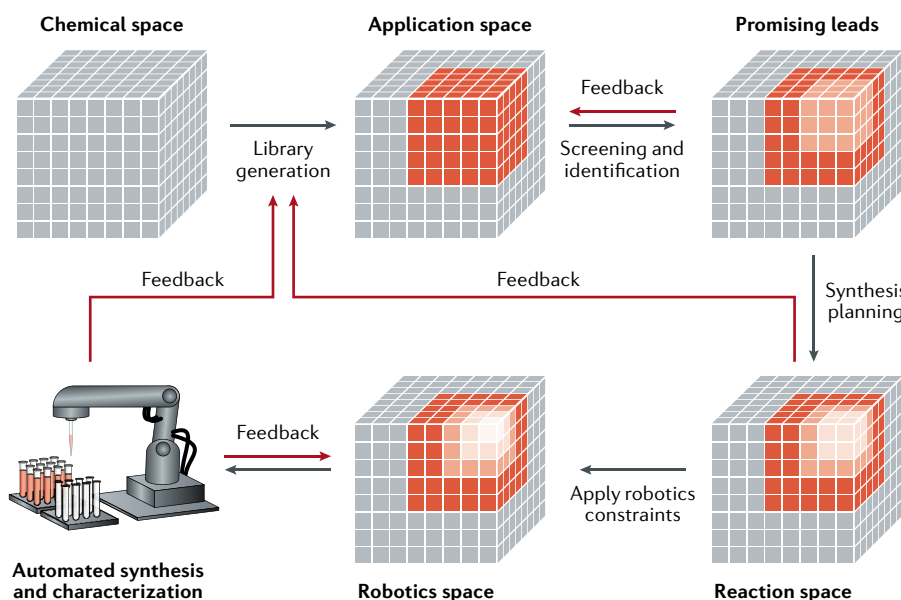
Fig. 2 | **Workflow of a closed-loop approach to autonomous materials discovery.** The procedure begins with identifying an application space of candidates for a given problem. The promising leads from this library are identified, potentially through computational screening, and are further narrowed by identifying the synthetically accessible molecules. Finally, the constraints of available robotics systems are taken into consideration before starting automated synthesis and characterization. Feedback from in situ experimentation is used to adjust the model, building the application space for the next iteration of this loop. Other feedback mechanisms at various stages of the loop aid in ensuring the candidates are compatible with all stages of the loop and reduce trial and error in the long term.

Virtual screening

High-throughput virtual screening (FIG. 3) already has a major role in the rational discovery of new functional materials, even without the automation of the other processes[174,175]. Organic materials problems, in particular, are highly amenable to high-throughput computation owing to the following three effects: first, organic materials are part of a very large chemical space; second, they have a large, robust set of descriptors; and finally, their performance can be reflected in a single descriptor. For example, reduction potentials can be computed from the energies of two different molecular states of charge, and the energy of a molecule is a thermodynamic quantity that electronic structure methods excel at computing. Similarly, for solar cells and organic light-emitting diodes (OLEDs), the key properties of the molecules are their frontier molecular orbital energies and the characteristics of their excited states, respectively.

We illustrate the current role of high-throughput virtual screening in organic materials discovery by reviewing three recent cases: OPVs[176,177], electrolytes for organic redox-flow batteries[168,170,178] and blue OLEDs[179]. These studies all involved the generation of molecular libraries containing on the order of 10⁶ molecules built out of

'building blocks' and showed that virtual screening can reduce the number of viable candidates to the range of tens to hundreds. In each of these cases, collaboration with experimental groups has been essential to maximizing the return of high-throughput virtual-screening applications. For the Clean Energy Project for the discovery of OPVs[176], the initial fragments that were used to generate the combinatorial library were recommended by the Bao group at Stanford University[180] ('human intuition' and 'library generation' blocks, FIG. 3). In the organic redox-flow-battery high-throughput virtual-screening studies, the decision to initially study quinones and to later study alloxazines was based on insight from synthetic organic chemists well versed in electrochemistry and the roles of these motifs in biological electron transport[168,170]. For the virtual screening of OLEDs, theorists received feedback from experimentalists by iterating through multiple batches of molecules, which often had some common attributes, for example, the backbone. The experimentalists voted on these molecules through an online voting system, and the best scoring and worst scoring molecules were then used to motivate the next generation of molecules to be screened. In each of these cases, experimental data were crucial for theorists to be able to calibrate

their models (this corresponds to the 'generate new molecules based on feedback' loop, FIG. 3). In the general case, some of the data can be obtained from the literature, but new types of materials may also require some new experiments to help with calibration (this corresponds to the 'model training and calibration' feedback loop, FIG. 3). It is important to note that the use of negative data collection is as important as positive feedback in training machine learning models[181].

High-throughput computational methods have also been successfully applied to inorganic materials[182–185]. The Materials Project[182], AFLOW[186] and the Open Quantum Materials Database[184] are leading examples of high-throughput databases of inorganic materials computed with density functional theory (DFT). These databases have enabled a large number of screening studies (the 'computational pipeline' block, FIG. 3) targeting the discovery of functional materials in applications ranging from batteries[187–190] and thermoelectrics[117,191] to catalysis[89,192]. These databases contain nearly all known inorganic crystalline compounds as well as a large number of hypothetical materials that span structural and chemical spaces unexplored by experiments. Current strategies that use such databases in the virtual screening of inorganic materials often start with the filtering of an initial large pool of candidate materials based on necessary conditions pertaining to the target application (for example, certain structural, thermodynamic and electronic requirements) to obtain a short-list of the most promising candidates.

Often, promising is not well defined, and multiple figures of merit[193] or selection rules[194] are used to grade potential performance. These criteria are grounded in fundamental properties, such as the bandgap, bulk modulus or defect formation energy — all of which can be calculated using off-the-shelf codes but at much higher computational costs than those of molecular systems. Further criteria that describe the viability of synthesis for a given material in select growth methods are beginning to emerge[195], whereas discernibility in characterization methodologies remains to be tackled. Taking the ease of synthesis into account in a materials screening approach enables optimization that is centred on not only how capable a potential material is but also how achievable and scalable its synthesis will be. Although collaborations may involve experimentalists who synthesize and evaluate the top candidates[196,197], it is not uncommon that the results of
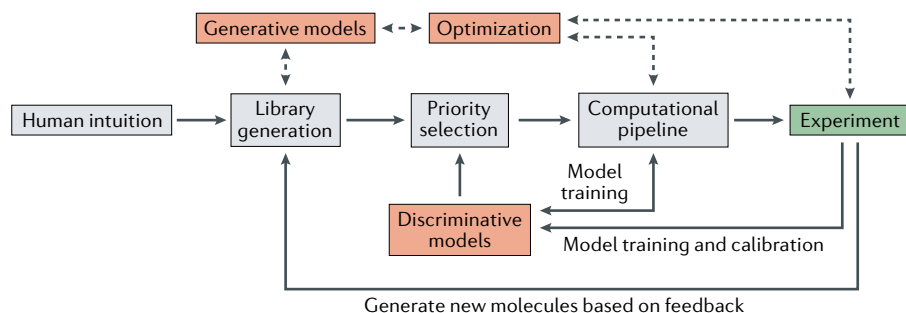
Fig. 3 | **State-of-the-art virtual screening: from human intuition to experimental verification.** The central path (in grey boxes) depicts the key stages in current implementations of the high-throughput virtual screening process. Initial libraries, or application spaces, are narrowed to a scale that can be subjected to a computational pipeline. This pipeline usually comprises quantum chemistry simulations, and the most promising candidates from this pipeline are experimentally tested. Recently, machine learning methods have been incorporated in the screening process at various levels (in red boxes). These include discriminative models to aid in both calibrating the computational pipeline and priority selection. In addition, generative models, used in conjunction with optimization and reinforcement learning, can propose candidates for experiment, obviating the need for combinatorial libraries.

virtual screening are disseminated to the community as recommendations for further experimental testing. Without strongly interrelated metrics, it is difficult to integrate experimental information on complex design criteria, such as cost and performance, within multicomponent applications across the energy materials landscape. This is also a key limitation in incorporating AI into the search process and optimizing compound properties rather than fundamental single structure properties, which are easy to calculate but often not representative of the end system owing to spatial and temporal scaling.

Recently, high-throughput virtual screening has become more integrated with machine learning processes. There are two avenues through which this has happened. First, machine learning regression algorithms can be used to calibrate theoretical models against experimental data in a more robust way than what can be achieved with simple linear regressions. Using methods of cheminformatics, such as molecular fingerprints and integrating with a Gaussian process, better fits can be obtained for calibration of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies[198]. Machine learning has also aided in the down selection of the initial candidates in the search for blue OLEDs. From an initial library with more than 1 million candidates, preliminary calculations on a subset of about 40,000 molecules enabled a neural network model (a discriminative model, FIG. 3) of thermally assisted delayed fluorescence rates (a key figure of merit) to be built[179]. This meant that only about one-quarter

of the full library was subjected to a full suite of more computationally expensive DFT calculations ('priority selection' block, FIG. 3). Machine learning and informatics techniques have also become one of the pillars for the computational discovery of inorganic materials, but progress in more representative descriptors and fingerprints for periodic crystals is essential for their more effective integration with virtual-screening strategies[118,199].

Other recent advances in machine learning have clear applications in high-throughput virtual screening. Studies of this type in the field of computational chemistry and materials are emerging on a daily basis. Many of these applications promise to accelerate virtual screening processes. For example, there is recent work that shows that deep learning is capable of obtaining DFT-accurate force fields that have the computational cost of classical force-field models (ANI-1)[40]. Such an advance may accelerate virtual screening procedures that would otherwise require the use of ab initio molecular dynamics simulations.

In the past year, there have been substantial developments in generative models for chemistry[51–55,200–202], which may change the way new molecules and materials are discovered. These models eliminate hand-coded combinatorial and rule-based generation of libraries for virtual screening. This is required because if the enumeration of molecules was pursued, the libraries for certain applications would become too large for even searching and hence also for the prediction of properties. New generative models could be used to generate focused, smaller virtual libraries that have shifted property distributions tailored

for targets[200,203] (FIG. 3). More importantly, these models can be used for inverse design (for example, inverse quantitative structure–activity relationship (QSAR)) by optimizing properties within the latent space and decoding back to molecules (see the linkage between 'generative models' and 'optimization', FIG. 3)[51]. Currently, most generative models use the simplified molecular-input line-entry system (SMILES; a text representation) as input, which does not have the ability to represent metal complexes, periodic materials or 3D geometry-dependent properties. In the near future, we expect to see more models that directly leverage molecular graphs and electron densities using the latest advances in generative model developments from other domains.

Virtual screening could add a step in the virtuous cycle described above by including the metric-driven modification of models used in relating input parameters to optimize target properties. One desired property that could be searched for is economic viability, for example, in the case of thermoelectric materials that can harvest electricity from heat flow or vice versa. However, the application of machine learning techniques to improve or even estimate production costs is still lacking.

Property optimization has been explored through the fitness landscape concept, which involves input–output mapping of the values of the variables used in modelling or designing candidate molecules to the value of the resultant objective property[204]. The OptiChem theorem has been presented as an explanation of the observed success of optimization in materials science with various combinations of objective parameters and experimental variables[205,206].

Overall, high-throughput virtual screening has matured to the point that it is a valuable tool in accelerating materials discovery, but its utility could be greatly enhanced by integrating its methodologies more closely with automated experimental exploration and feedback of the molecules and materials that it proposes for development.

Autonomous synthesis planning
Synthesis planning for inorganic materials. Synthesis planning for bulk inorganic materials is still in its infancy relative to molecular and organic systems. Although centuries of metallurgy and ceramography have yielded substantial insight into synthesis and processing, most of this understanding is still empirical. Predictive synthesis has relied on translating empirical concepts

into computable analogues that are centred on finding local and global minima of parameterized free energies[207,208]. For example, phase diagrams depict thermodynamic stability as a function of external constraints and can be constructed using convex hull methods on computed formation energies[209]. Legendre transformations convert formation energies into other forms relevant to different areas of synthetic science, including Pourbaix diagrams[210] for aqueous stability and Ellingham diagrams[211] for stability in oxygen-rich environments. Such phase diagrams represent different parametric slices of the global free-energy space. Free energies themselves can be computed from first principles or semi-empirical fits, such as those used by the CALPHAD community[212], to understand and improve metallurgical systems.

Real-world materials, however, contain both chemical and structural heterogeneities that are often closely linked with the dynamics of synthesis. The nucleation and growth of grain boundaries, dislocations, defects, segregations and many other features may completely redefine the properties of the material relative to the simple crystalline bulk phase[213]. Although these phenomena are becoming more tractable[214–216] in theoretical simulations, predictive modelling to control any of these macroscopic features is still very difficult and is typically restricted to select high-value systems. Therefore, notable advances in both theoretical approaches and the aggregation of experimental data with respect to these properties will likely be necessary to fully realize the power of AI and machine learning tools for inorganic materials synthesis prediction.

The formation and reformation of crystalline phases are often not as simple as the intermediates and transition states in organic synthesis. However, the key concepts that are needed to model these transitions, including diffusion[217] and structural deformation[218,219], are gaining increasing attention from the theoretical community. In addition, new approaches to understanding metastability enable the energetic upper-bound of phase transitions between material polymorphs to be estimated[220], and a growing body of classical force-field data[221] might be leveraged to approximate kinetics in a similar way.

Synthesis planning for organic molecules.
Modern virtual high-throughput screening of organic molecules joins combinatorial or generative models together with in silico predictions and cheminformatics to facilitate the efficient and targeted exploration of chemical space[51,198]. As the best-performing candidates are not guaranteed to be the most feasible for actual synthesis, typically, human experts must select the most promising candidates and invest substantial amounts of time in synthesis planning and device fabrication. Thus, automating the prediction of chemical synthesis is one of the central pillars in a fully autonomous search for high-performance materials discovery. Below, we review traditional rule-based expert systems suited for retrosynthesis as well as more recent developments in machine learning approaches to predict reaction outcomes[222].

Computer-aided synthesis planning has existed for more than 40 years, dating back to the formalization of retrosynthesis in the seminal work by Corey[223]. The strategy of retrosynthesis is to leverage accessible chemical reactions to trace the target molecule back to commercially available or simpler to synthesize starting materials. Formally, the reaction network of products that are linked to substrates via reactions can be represented as a bipartite graph. Synthesis planning is thereby reduced to a search of possible routes that connect the target molecule to suitable substrates. Depending on the depth of the search — that is, the number of reaction steps — there is typically a large set of potential synthetic routes. Therefore, autonomous retrosynthesis requires the implementation of search strategies to avoid a combinatorial explosion of possibilities. Heuristics are introduced to rank the different synthetic routes on the basis of specific criteria, such as the cost of the substrates or number of synthetic steps.

Corey and Wipke presented the first computer-aided synthesis design software, called Organic Chemical Simulation of Synthesis (OCSS)[224], in 1969. This was quickly followed by LHASA[225,226], which implements search strategies based on functional groups or structural features to find the most reasonable synthetic route. Various computer-aided programs were implemented until the late 1990s, such as CAMEO[227], EROS[228], SOPHIA[229] and SYNCHEM[230], which all depend on expert rules that needed to be hand-coded and required substantial human effort. Therefore, the mapped chemical space was quite small, and the early software packages were of limited value for expert chemists.

Advances in computational resources and computer science, as well as accessible databases containing large numbers of known reactions and molecules (for example, Reaxys, InfoChem and ChemPlanner), have led to a breakthrough in automated synthesis planning. Thus, several commercial products are now available[231]. Their specific implementations differ, but the general concept is very similar. Typically, databases are scanned and reaction rules are extracted automatically[232], followed by unsupervised learning to cluster similar reactions into groups, which finally allows for rule generalization.

Automated rule extraction begins with identifying the reaction core, which is determined by the atoms involved in bond formation or breaking during the reaction. As reactions are significantly influenced by the chemical environment, it is of importance to extend the reaction core to include neighbouring atoms or functional groups to encode structural information[233]. The drawbacks of the current approaches are the inability to properly account for chemical context and the stereochemistry or regiochemistry of a reaction. Therefore, rule curation becomes indispensable and can be done by either verifying the extracted reaction templates against known databases (for example, the Beilstein database) or through manual correction by expert chemists. For example, the retrosynthesis module of Chematica[234], Syntaurus, contains 20,000 manually coded rules that were implemented by expert organic chemists.

The extracted rules together with the substrates form a large reaction network. Typically, breadth-first search-like algorithms are used to expand possible synthetic routes[233,235]. The Syntaurus package implements an elegant algorithm that simultaneously explores potential synthetic routes and uses scoring functions to estimate the value of the current synthetic position. The latter allows the search algorithm to make more informed decisions about which synthetic pathways will be expanded and which ones are considered inefficient and thus will be terminated. The search along a certain reaction pathway is continued only as long as it minimizes a predefined cost function.

Although rule-based expert systems have made a considerable leap forward in recent years, they suffer from their inability to fully include chemical context. Therefore, manual rule curation is required, and reaction templates need to be augmented by additional information, which involves the time-consuming manual encoding of explicit lists specifying protected or incompatible functional groups. Additionally, the prediction of unknown reactions outside the chemical space from which the rules were extracted is a major challenge for rule-based expert systems.

The shortcomings of rule-based expert systems have provoked the idea to leverage AI approaches to reaction prediction[236] (FIG. 4). Several recent studies[237–239] have involved training neural networks as a classifier to predict reaction outcomes on a database of known reactions. For example, the reaction outcomes for 16 simple reactions of alkene and alkyl halides were successfully predicted using this approach[237]. Once trained, a neural network provides a ranking of potential reaction outcomes for a specific set of reaction rules and suitable chemical descriptors, such as neural[240] or Morgan fingerprints (implemented in cheminformatics software, such as RDKit). The capability of these neural-network-based reaction prediction algorithms is determined by the quality and size of the chemical space that is represented by the underlying database. To be able to predict outcomes for a wide range of chemical reactions, it is necessary to use a large and diverse database in training. In one study, millions of reactions were taken from the Reaxys database to train a neural network[238]. For a set of 8,720 automatically extracted reaction templates, the highway network returned the correct result in 78% of the specified test cases, and in 98% of the cases, the network ranked the correct reaction within its top ten suggestions. Similarly, reaction templates taken from granted US patents were applied to a set of reactants, generating a pool of plausible products[239].

One of the greatest challenges for machine learning approaches lies in the unavailability of failed reaction data, as publications are heavily biased towards positive results. However, similar to humans, learning from failure is a key component of machine learning chemical reactions[241]. In the context of materials discovery, information gained from unsuccessful reactions was demonstrated to be crucial for predicting reaction outcomes for the crystallization of templated vanadium selenites[181]. In this study, a vast number of unreported failed reactions were extracted from laboratory notebooks and leveraged a support vector machine model to establish a reverse-derived decision tree. The decision tree provided simple guidelines that helped chemists in decision making during a synthesis process.

In tandem with automated rule-based synthesis planning, machine-learning-based rankings of most likely reaction outcomes facilitate a faster and more targeted planning of synthetic routes. Such a combined approach could pave the way for next-generation synthesis planning software and would be of particular relevance for planning new reaction pathways for novel products. Reaction prediction could be applied to rule out unviable reaction pathways as well as to display a ranking of the most realistic routes for actual synthesis. For molecules that have no predicted synthetic routes, feedback should be given to the virtual-screening system (FIG. 2).

Once a synthesis tree is generated by the autonomous synthesis planning module (FIG. 4), the instructions need to be translated into a machine-readable language to start the reactions along the synthesis loop.

Automated chemical synthesis

High-throughput chemistry, often referred to as automated chemistry, was pioneered by the pharmaceutical industry in the search for efficient screening of their application spaces and for the cost-effective synthesis of new organic compounds[242,243]. In the mid to late 1990s, the first automated laboratory emerged in the field of peptide chemistry[244]. Many pharmaceutical companies, including Eli Lilly and Company, Merck and Company and Aventis Pharma, now use integrated automated chemical synthesis laboratories for an efficient, although combinatorial, search of their application spaces in the field of medicinal chemistry[243,245,246]. Compared with traditional experimental procedures, automated chemistry enables parallelization of many experiments while requiring fewer resources per experiment[245]. Machine-assisted equipment also drastically reduces the exposure of humans to hazards, such as toxic solvents and explosives[247]. Therefore, fully automated laboratories powered by AI will be a major break through in the approach to materials discovery.

Autonomous laboratories are expected to expand their application spaces to new fields, such as PVs and materials for energy storage. To this end, flexible and modular autonomous systems are required to cover a broad range of chemical space and need to facilitate multiple sets of reactions and multivariant environments (for example, different temperatures, pressures and solvents)[18,247,248]. Several automation strategies have emerged in recent years, including flow chemistry[249,250],
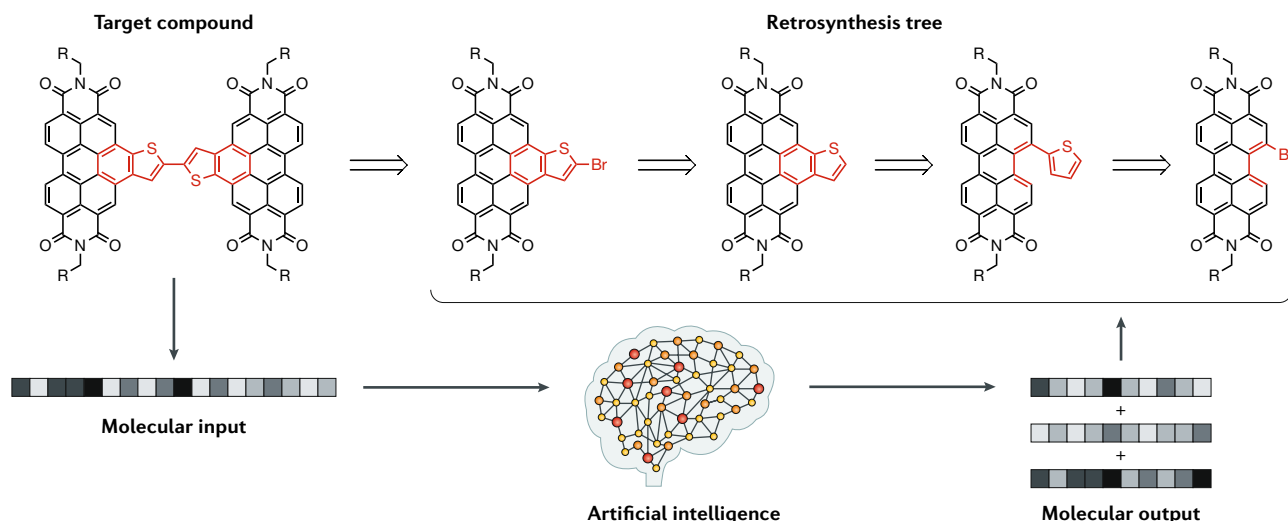

Fig. 4 | **General concept for the automated generation of retrosynthesis trees.** From a promising lead for organic photovoltaics[292] (left), descriptors (depicted as fingerprints for illustrative purposes) are used as input for artificial intelligence algorithms (for example, neural networks), which output the descriptors for molecules in the retrosynthesis tree. The sections of the molecules involved in each retrosynthetic step are highlighted in red.

microfluidic systems[251] and nanomole-scale batch miniaturization, for natural product synthesis[252]. The reaction space typically includes coupling reactions, such as Suzuki–Miyaura, Stille, Heck, Sonogashira, Negishi, Friedel–Crafts, Wittig, amination and direct arylation reactions[253–257]. All these coupling reactions can be viewed as chemical Lego, with small building blocks being combined in the presence of catalytic materials. A promising automated synthesis device was recently developed[258–260]. In this approach, 12 boronate-ester building blocks for small organic molecules were synthesized, creating an accessible chemical space for automated synthesis that covers 75% of polyene natural products. Recently, the same setup was extended to 14 new classes of small molecules, including complex natural products containing macrocyclic or polycyclic frameworks[261]. Estimates show that about 5,000 building blocks would be sufficient to synthesize 70–75% of the nearly 260,000 small-molecule natural products[259]. Another automated synthesis, purification and testing platform was also recently developed for small molecules with biological relevance[262].

The autonomous synthesis of molecules requires the chemical reactions and characterization of the products to be linked through feedback loops at the hardware level. This raises technical and theoretical challenges at various stages. The first challenge is the design of the hardware assembly. Traditional reaction schemes and characterization techniques need to be modified and adapted from batch to flow chemistry, which has emerged as a leading strategy to reach automation[18]. The second challenge relates to software architecture. Automated machinery would generate large quantities of data, which would need to be analysed. This in turn requires long pipelines and the development of workflow platforms. Data would be generated by analytical tools and would need to be stored in a database for further optimization of the reaction conditions. Notably, optimization of the reactions will be done based on the concentration of catalysts, types of ligands, concentrations of additive salts and types of solvents. The target of the optimization procedure could include the yield of the reaction, the presence of by-products that may be toxic or contaminate the main product, or the production cost. Outputs from the different modules should be made compatible with the next step along the workflow and, importantly, the vast number of tasks that must run in a specific sequence. Overall, orchestration

of the aforementioned tools, at both the hardware and software level, makes the construction of workflows (for example, Awesome Pipeline) challenging[183].

There are also notable automated chemistry efforts outside of organic chemistry. Combinatorial thin-film deposition is emerging as one such method[263]. Its development dates back to the mid-1990s[264], with commercialization by companies such as Symyx Technologies, PVD Products, Neocera and Intermolecular spanning a wide application space. The systems developed by these companies can deposit whole libraries of compositions, starting from elemental or even compound sources[265–272]. This enables searches for individual phases, disordered structures, solid solutions, multilayered heterostructures, super-lattices and other discoveries. The resulting multivariant compositions are well suited to a variety of characterization suites that provide information on inherent structure (for example, microstructure, composition, disorder and distortions) as well as functionality (for example, electrical, thermal, magnetic and optical properties). The ease of characterization also creates an opportunity for an integrated system that couples theory, synthesis and characterization, but although this has been achieved, it is not commonly implemented.

So far, the use of automated chemistry is mostly a trial-and-error approach. However, the community has started exploring the next generation of machine-assisted laboratories for materials discovery (for example, Dial-a-Molecule), in which automation is present at every stage of the development process[241]. The approach we propose here goes a step beyond automation and adds AI for a more rational discovery route that uses feedback collected through experimental measurements and characterization. Such an approach will increase the productivity, reduce the cost and enable new types of experiments[26], which are key components for innovation. Increases in discovery rate must go hand in hand with an equally efficient characterization of the system.

High-throughput characterization
Efficient materials and device characterization techniques are crucial elements in the autonomous material discovery approach outlined in FIG. 2. These techniques are used for the rapid automated characterization of chemical composition, materials structure, physical properties and device functionality. Characterization should

be naturally integrated with the synthetic step, and compatible workflows with a feedback procedure, preferably at a hardware level, should be used. The characterization step contains three different stages (FIG. 5).
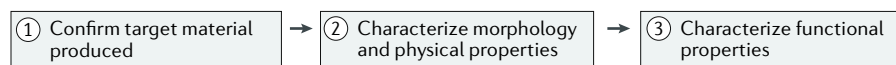
The first, and the most common, analytical step is to confirm that the target chemicals and/or materials are produced. This step usually involves mass spectrometry, spin resonance spectroscopy and optical spectroscopy techniques, and it is closely connected with the materials synthesis. For example, several analytical techniques have been adapted to automatic flow chemistry[19]. These include high-throughput NMR spectroscopy[273,274], which enables monitoring of the reaction process in real time[261], and optical methods, such as UV–visible, Fourier-transform infrared (FTIR) and Raman spectroscopy, which provide information on the structures. Notably, on-line FTIR spectroscopy is one of the most efficient and reliable tools for monitoring and controlling the progress of flow processes[275,276]. High-throughput, spectrally resolved photoluminescence scanning has recently been suggested as a key method for the characterization of optoelectronic materials[112]. However, automatic materials discovery will still greatly benefit from a broader set of analytical tools for fingerprinting of the synthesized compounds. It is also important to diversify the autonomous platforms into which these techniques are implemented. Although analytical tools can be applied to various materials independently of their actual applications, particular chemical reactions may be easier, for example, using a batch rather than flow setup, despite the advantages of flow reactions outlined in the previous section.

The second step involves the characterization of the physical properties, morphologies, defects and interfaces of the functional materials. At this step, high-throughput techniques developed within the Materials Genome Initiative[13,14] can be adapted in a straightforward manner. For example, the high-throughput devices used for measuring Seebeck coefficients[277] can be used for the optimization of thermoelectric modules. These devices enable films to be probed with combinatorial material compositions. Robotic systems for probing film properties are used in both industry and academia. Mechanical properties of coatings were analysed using robotic systems at The Dow Chemical Company, where the characterization methods included microindentation, measurements of impact resistance and measurement of friction
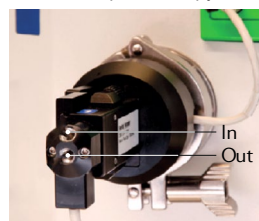
Characterization toolkit

Optical spectroscopy:	Microscopy:	Spin resonance:	Electrical characterization:	Mechanical characterization
• UV–vis, IR absorption	• optical, IR, X-ray	• NMR, EPR	• transport	
• fluorescence	• SEM, TEM, STM		• magnetoresistance	Mass spectrometry
• raman scattering	• AFM		• capacitance	

Characterization step

① Confirm target material produced → ② Characterize morphology and physical properties → ③ Characterize functional properties
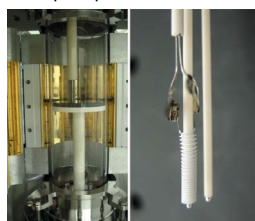
Example

On-line IR spectroscopy · Transport probe · Roll-to-roll testing

Fig. 5 | **High-throughput characterization of materials.** Characterization techniques that operate at several scales (upper part). Characterization has three different steps (middle part). Initial characterization focuses on verification, using analytical tools, of the synthesis of the target (step 1). Subsequent characterization focuses on the morphology and properties of the material (step 2). Finally, the functional performance of the device is characterized (step 3). Examples of these three stages of characterization (lower part) include on-line IR spectroscopy[275] for target material verification, a high-throughput probe for measuring the transport properties of thermoelectric materials[293] and automatic testing of polymer solar cells integrated into a roll-to-roll processing method[280]. AFM, atomic force microscopy; EPR, electron paramagnetic resonance; SEM, scanning electron microscopy; STM, scanning tunnelling microscopy; TEM, transmission electron microscopy; vis, visible. Lower left panel is reproduced with permission from REF.[275], American Chemical Society. Images in the lower middle panel are reproduced with permission of J. Martin. Lower right panel is reproduced with permission from REF.[280], American Chemical Society.

coefficients[278]. The surface topography of the films was also characterized using white-light interferometry. An automatic high-throughput analysis of OPV films for electrical defects has been conducted using infrared detection combined with an algorithmic segmentation of PV cells and defects[279]. These defects lead to short-circuit currents, reduce the performance of the cells and are identified by local heat dissipation.

The third and final step involves the characterization of the functional properties of the devices and also addresses stability issues, for example, using accelerated lifetime tests. This final step is probably the most challenging in high-throughput characterization. A functional device typically comprises several layers that have to be processed on top of each other without affecting the quality of the underlying layers. This is demanding for any robot-based production method, and it is most challenging for the correct interpretation of device performance. As device performance is integrative, recording, for example, the efficiency of a solar cell

will not give conclusions as to whether the semiconductor, the interface or the whole stack performed to expectations. Testing methods at this step are most specific for particular applications.

High-throughput device characterization has been implemented in the roll-to-roll processing of organic solar cells, for which the donor:acceptor ratio and layer thickness parameters were screened in assembled devices[280]. More recent developments are starting to integrate high-throughput instrumentation into synchrotron beam lines to measure in situ the properties of materials during synthesis or films during drying[281]. High-power illumination sources with the intensity of several hundred suns are becoming available to rapidly test novel materials under harsh conditions in the laboratory. A certified 1,000 hour solar cell test could be completed within hours on such a setup. Further activities need to focus not only on characterizing materials but also on measuring properties of whole devices, for which transient electrical and optical spectroscopy methods are

required. Hyperspectral imaging methods in combination with imaging analysis and machine-learning-based image recognition are expected to enable large numbers of materials and devices to be screened within a short time period and with unprecedented precision. Automated materials discovery will greatly benefit from all these characterization techniques, provided that they are integrated at both the hardware and software levels with the other elements in the cycle of materials discovery and innovation. However, a generalized platform has yet to be designed and developed.

Autonomous experimentation
To fully exploit the advances in autonomous robotics, machine learning, high-throughput virtual screening, combinatorial methods and in situ or in operando characterization, we must close the loop in the research process. This means that humans must partner with autonomous research robots to design experimental campaigns and that the research robots perform experiments, analyse the results, update our understanding and then use AI and machine learning to design new experiments optimized to the research goals, thus completing one experimental loop (FIG. 6). Several groups have demonstrated this closed-loop approach in a range of applications, including carbon nanotubes, Bose–Einstein condensates, alloys, substituted functional organic molecules, oil droplets and the search for new chemical reactions[23–26,57,282–284]. A notable example is the development of the Autonomous Research System, ARES[23], to optimize the synthesis of carbon nanotubes. ARES learned to grow carbon nanotubes autonomously using automated experimentation, in situ characterization and machine learning for optimal experimental design. The AI planner for this iteration of ARES used a random forest representation of the prior experiments using Lockheed Martin's Nanotechnology Materials Data Mining, Modelling and Management (NMD-M3) software[285]. The planner provided input conditions (for example, gas composition, temperature and pressure) that were expected to achieve the growth rate supplied by the human researcher. Raman spectroscopy was performed in situ during each experiment to characterize and quantify the growth rate for the experimental conditions provided by the planner. The database was then updated with the experimental input conditions and the resultant growth rate.
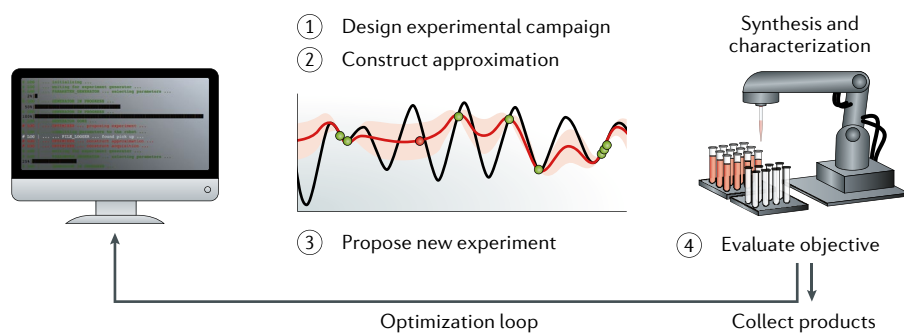
Fig. 6 | **Autonomous experimentation procedures.** Autonomous experimentation bridges computational power and robotics solutions to create a virtuous cycle. The approximation function (red) models how a particular set of experimental conditions will perform in terms of, for example, yield and reaction time. Once constructed, this function is used to propose the next experiment. The model is updated and moves towards maximizing the performance of the reaction based on the predefined conditions.

It is important to distinguish between automated systems, which are able to perform repetitive, pre-planned tasks with human direction, and autonomous systems, which are able to adapt appropriately to new information to drive towards human-defined goals without human intervention. Autonomous systems have the advantage of exploiting information from the most recent iteration to perform optimal experimental design of the next experiments. The gain in experimental efficiency by avoiding uninformative experiments and by intentionally probing the most informative and productive experimental conditions leads to exponential gains in the speed of research progress. It is also important to distinguish the closed-loop approach from high-throughput and/or combinatorial methods, which lack on-the-fly adaptability. In the future, autonomous experimentation systems will incorporate autonomous hypothesis generation and testing, integrated modelling and simulation, and data sciences to further increase the rate of research. Indeed, in the future, there may be a "Moore's Law for the speed of research" (REF.[286]) in which the rate of research climbs exponentially over the next few decades.

Conclusions and outlook

Autonomous materials discovery in the field of clean energy production powered by AI provides a universal platform for the emergence of novel solutions towards a low-carbon economy. Although they are not yet ready, fully integrated autonomous platforms are at our fingertips, reachable in the next 5–10 years. However, this would require combined multidisciplinary and multinational efforts between research institutions, industry, private investors, and public and governmental organizations. It is

important that the broader public is informed and supportive of the challenges and, most importantly, of the opportunities associated with the revolution in energy innovation. Furthermore, scientific information needs to be reported and widely disseminated in a more comprehensible form.

Advances in high-performance and low-cost materials will be essential to power and operate the low-carbon global economy. For example, today's sensor and transmitter materials consume too much energy to enable widespread interconnectivity between smart devices[287]. Hence, improved, low-energy semiconductor technology is imperative to implement a scenario in which the Internet of Things becomes a widespread reality, with billions of devices, tools and equipment interconnected via the Internet.

From a technical perspective, accelerated materials innovation requires stronger synergy between the experimental and theoretical components of the platform. From a theoretical point of view, the data generation and accessibility that are necessary for enabling robust AI algorithms are among the greatest challenges. From an experimental point of view, AI will support experimentalists in their strategic goal to find materials with specific desired properties more quickly and more accurately. Currently, the bottleneck is the experimental synthesis, characterization and testing of theoretically proposed materials. If the above approach is successfully implemented, the bottleneck will move to AI.

We have the opportunity to enable a world powered by sustainable energy if we leverage technological innovation in materials science, chemistry and computer science. In the short term, the role of academia is to demonstrate that autonomous workflows, integrating high-throughput computations and experiments, can lead to accelerated novel

materials discovery. In the long term, these methods will be adopted by industry, and the resulting materials will compete for successful commercialization. The semiconductor industry has been highly successful in both planning, through the International Technology Roadmap for Semiconductors, and following through in achieving their outlined goals. Similar strategies should be applied in materials discovery for clean energy technologies in the future.

Daniel P. Tabor [iD] [1], *Loïc M. Roch* [iD] [1], *Semion K. Saikin* [iD] [1], *Christoph Kreisbeck* [1], *Dennis Sheberla* [iD] [1], *Joseph H. Montoya* [iD] [2], *Shyam Dwaraknath* [2], *Muratahan Aykol* [iD] [3], *Carlos Ortiz* [4], *Hermann Tribukait* [5], *Carlos Amador-Bedolla* [iD] [6], *Christoph J. Brabec* [iD] [7,8], *Benji Maruyama* [9], *Kristin A. Persson* [iD] [2,10] *and Alán Aspuru-Guzik* [iD] [1]*

[1]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA.*

[2]*Energy Storage and Distributed Resources Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.*

[3]*Toyota Research Institute, Los Altos, CA, USA.*

[4]*Secretaría de Energía, México Del Valle, Mexico City, Mexico.*

[5]*Fondo de Sustentabilidad Energética, Mexico City, Mexico.*

[6]*Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, Mexico.*

[7]*Department of Materials Science, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Erlangen, Germany.*

[8]*Renewable Energy Division, ZAE Bayern, Erlangen, Germany.*

[9]*Air Force Research Laboratory, Materials and Manufacturing Directorate, Wright–Patterson Air Force Base, Dayton, OH, USA.*

[10]*Department of Materials Science, University of California Berkeley, Berkeley, CA, USA.*

*e-mail: aspuru@chemistry.harvard.edu

1. Dunn, B., Kamath, H. & Tarascon, J.-M. Electrical energy storage for the grid: a battery of choices. *Science* **334**, 928–935 (2011).
2. She, X., Huang, A. Q. & Burgos, R. Review of solid-state transformer technologies and their application in power distribution systems. *IEEE J. Emerg. Sel. Top. Power Electron.* **1**, 186–198 (2013).
3. Mahlia, T. M. I., Saktisahdan, T. J., Jannifar, A., Hasan, M. H. & Matseelar, H. S. C. A review of available methods and development on energy storage; technology update. *Renew. Sustain. Energy Rev.* **33**, 532–545 (2014).
4. Chabot, V. et al. A review of graphene and graphene oxide sponge: material synthesis and applications to energy and the environment. *Energy Environ. Sci.* **7**, 1564–1596 (2014).
5. Ferreira, A. D. B., Nóvoa, P. R. & Marques, A. T. Multifunctional material systems: a state-of-the-art review. *Compos. Struct.* **151**, 3–35 (2016).
6. Werber, J. R., Osuji, C. O. & Elimelech, M. Materials for next-generation desalination and water purification membranes. *Nat. Rev. Mater.* **1**, 16018 (2016).
7. Maine, E. & Garnsey, E. Commercializing generic technology: the case of advanced materials ventures. *Res. Policy* **35**, 375–393 (2006).
8. Linton, J. D. & Walsh, S. T. From bench to business. *Nat. Mater.* **2**, 287–289 (2003).
9. Sabatier, M. & Chollet, B. Is there a first mover advantage in science? Pioneering behavior and

scientific production in nanotechnology. *Res. Policy* **46**, 522–533 (2017).

10. Jackson, R. B. in *New U.S. Leadership, Next Steps on Climate Change* (ed. Hayes, D. J.) 129–135 (Stanford Woods Institute for the Environment, Stanford, CA, USA, 2016).

11. Georgeson, L., Maslin, M. & Poessinouw, M. Clean up energy innovation. *Nature* **538**, 27–29 (2016).

12. Bernstein, A. et al. Renewables need a grand-challenge strategy. *Nature* **538**, 30 (2016).

13. [No authors listed.] The first five years of the materials genome initiative: accomplishments and technical highlights. *Materials Genome Initiative* https://www.mgi.gov/sites/default/files/documents/mgi-accomplishments-at-5-years-august-2016.pdf (2016).

14. Green, M. L. et al. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. *Appl. Phys. Rev.* **4**, 011105 (2017).

15. UNFCCC. *Adoption of the Paris Agreement. Report No. FCCC/CP/2015/L.9/Rev.1* (UNFCCC, 2015).

16. Northrop, E., Biru, H., Lima, S., Bouyé, M. & Song, R. Examining the alignment between the intended nationally determined contributions and sustainable development goals. *World Resources Institute* https://www.wri.org/sites/default/files/WRI_INDCs_v5.pdf (2016).

17. Knight, W. The dark secret at the heart of AI. *MIT Technology Review* https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (2017).

18. Ley, S. V., Fitzpatrick, D. E., Ingham, R. J. & Myers, R. M. Organic synthesis: march of the machines. *Angew. Chem. Int. Ed.* **54**, 3449–3464 (2015).

19. Schrage, M. 4 Models for using AI to make decisions. *Harvard Business Review* https://hbr.org/2017/01/4-models-for-using-ai-to-make-decisions (2017).

20. Geysen, H. M., Meloen, R. H. & Barteling, S. J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl Acad. Sci. USA* **81**, 3998–4002 (1984).

21. Doyel, P. M. Combinatorial chemistry in the discovery and development of drugs. *J. Chem. Technol. Biotechnol.* **64**, 317–324 (1995).

22. Borman, S. Combinatorial chemistry. *Chem. Eng. News* **76**, 47–67 (1998).

23. Nikolaev, P. et al. Autonomy in materials research: a case study in carbon nanotube growth. *Comput. Mater.* **2**, 16031 (2016).

24. Wigley, P. B. et al. Fast machine-learning online optimization of ultra-cold-atom experiments. *Sci. Rep.* **6**, 25890 (2016).

25. Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, 11241 (2016).

26. Houben, C. & Lapkin, A. A. Automatic discovery and optimization of chemical processes. *Curr. Opin. Chem. Eng.* **9**, 1–7 (2015).

27. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

28. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & de Freitas, N. Taking the human out of the loop: a review of Bayesian optimization. *Proc. IEEE* **104**, 148–175 (2016).

29. Allen, K. How a Toronto professor's research revolutionized artificial intelligence. *thestar.com* https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html (2015).

30. Gibney, E. Google AI algorithm masters ancient game of Go. *Nature* **529**, 445–446 (2016).

31. Cisco Public. Encrypted traffic analytics. *Cisco* https://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise-networks/enterprise-network-security/nb-09-encrytd-traf-anlytcs-wp-cte-en.pdf (2018).

32. Basuchoudhary, A., Bang, J. T. & Sen, T. *Machine-Learning Techniques in Economics*. (Springer, Berlin, 2017).

33. Mullainathan, S. & Spiess, J. Machine learning: an applied econometric approach. *J. Econ. Perspect.* **31**, 87–106 (2017).

34. Rao, A. Digital twins beyond the industrials. *PWC* http://usblogs.pwc.com/emerging-technology/digital-twins/ (2017).

35. Magoulas, G. D. & Prentza, A. in *Machine Learning and Its Applications. ACAI 1999. Lecture Notes in Computer Science* Vol 2049 (eds Paliouras, G., Karkaletsis, V. & Spyropoulos, C. D.) 300–307 (Springer, Berlin, 2001).

36. Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C. & Ng, A. Y. Cardiologist-level arrhythmia detection with convolutional neural networks. Preprint at *arXiv*, 1707.01836 (2017).

37. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).

38. Shen, G., Horikawa, T., Majima, K. & Kamitani, Y. Deep image reconstruction from human brain activity. Preprint at bioRxiv, 240317 (2017).

39. Goh, G. B., Hodas, N. O. & Vishnu, A. Deep learning for computational chemistry. *J. Comput. Chem.* **38**, 1291–1307 (2017).

40. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).

41. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. *Neural message passing for quantum chemistry*. Preprint at *arXiv*, 1704.01212 (2017).

42. Matlock, M. K., Dang, N. L. & Swamidass, S. J. Learning a local-variable model of aromatic and conjugated systems. *ACS Cent. Sci.* **4**, 52–62 (2018).

43. Jiménez, J., Škalic, M., Martinez-Rosell, G. & De Fabritiis, G. K$_{DEEP}$: protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J. Chem. Inf. Model.* **58**, 287–296 (2018).

44. Wang, H. & Yeung, D.-Y. Towards Bayesian deep learning: a framework and some existing methods. *IEEE Trans Knowl. Data Eng* **28**, 3395–3408 (2016).

45. Ehsan Abbasnejad, M., Shi, Q., Abbasnejad, I., van den Hengel, A. & Dick, A. Bayesian conditional generative adversarial networks. Preprint at *arXiv*, 1706.05477 (2017).

46. Häse, F., Roch, L. M., Kreisbeck, C. & Aspuru-Guzik, A. PHOENICS: a universal deep Bayesian optimizer. Preprint at *arXiv*, 1801.01469 (2018).

47. Hansen, K. et al. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theor. Comput.* **9**, 3404–3419 (2013).

48. Brockherde, F. et al. Bypassing the Kohn–Sham equations with machine learning. *Nat. Commun.* **8**, 872 (2017).

49. Li, Z., Kermode, J. R. & De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).

50. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet — a deep learning architecture for molecules and materials. Preprint at *arXiv*, 1712.06113 (2017).

51. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).

52. Blaschke, T., Olivecrona, M., Engkvist, O., Bajorath, J. & Chen, H. Application of generative autoencoder in de novo molecular design. *Mol. Inf.* **37**, 1700123 (2018).

53. Sánchez-Lengeling, B., Outeiral, C., Guimaraes, G. L. & Aspuru-Guzik, A. Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). Preprint at *ChemRxiv* https://doi.org/10.26434/chemrxiv.5309668.v3 (2017).

54. Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A. & Zhavoronkov, A. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Mol. Pharm.* **14**, 3098–3104 (2017).

55. Grover, A., Dhar, M. & Ermon, S. Flow-GAN: combining maximum likelihood and adversarial learning in generative models. Preprint at *arXiv*, 1705.08868 (2017).

56. Duros, V. et al. Human versus robots in the discovery and crystallization of gigantic polyoxometalates. *Angew. Chem. Int. Ed.* **56**, 10815–10820 (2017).

57. Zhou, Z., Li, X. & Zare, R. N. Optimizing chemical reactions with deep reinforcement learning. *ACS Cent. Sci.* **3**, 1337–1344 (2017).

58. King, R. D. et al. The automation of science. *Science* **324**, 85–89 (2009).

59. Trancik, J. E. Renewable energy: back the renewables boom. *Nature* **507**, 300–302 (2014).

60. Naims, H. Economics of carbon dioxide capture and utilization — a supply and demand perspective. *Environ. Sci. Pollut. Res.* **23**, 22226–22241 (2016).

61. Muratori, M. et al. Carbon capture and storage across fuels and sectors in energy system transformation pathways. *Int. J. Greenhouse Gas Control* **57**, 34–41 (2017).

62. Tzimas, E. et al. CO$_2$ utilisation today: report 2017. *DepositOnce* https://doi.org/10.14279/depositonce-5806 (2017).

63. Kuhl, K. P., Cave, E. R., Abram, D. N. & Jaramillo, T. F. New insights into the electrochemical reduction of carbon dioxide on metallic copper surfaces. *Energy Environ. Sci.* **5**, 7050–7059 (2012).

64. Kuhl, K. P. et al. Electrocatalytic conversion of carbon dioxide to methane and methanol on transition metal surfaces. *J. Am. Chem. Soc.* **136**, 14107–14113 (2014).

65. Roberts, F. S., Kuhl, K. P. & Nilsson, A. High selectivity for ethylene from carbon dioxide reduction over copper nanocube electrocatalysts. *Angew. Chem. Int. Ed.* **127**, 5268–5271 (2015).

66. Reymond, H., Vitas, S., Vernuccio, S. & von Rohr, P. R. Reaction process of resin-catalyzed methyl formate hydrolysis in biphasic continuous flow. *Ind. Eng. Chem. Res.* **56**, 1439–1449 (2017).

67. Behrens, M. Heterogeneous catalysis of CO$_2$ conversion to methanol on copper surfaces. *Angew. Chem. Int. Ed.* **53**, 12022–12024 (2014).

68. Kattel, S., Ramírez, P. J., Chen, J. G., Rodriguez, J. A. & Liu, P. Active sites for CO$_2$ hydrogenation to methanol on Cu/ZnO catalysts. *Science* **355**, 1296–1299 (2017).

69. U.S. Energy Information Agency. *Manufacturing Energy Consumption Survey (MECS) 2014* (U.S. Energy Information Agency, 2014).

70. Reymond, H., Amado-Blanco, V., Lauper, A. & Rudolf von Rohr, P. Interplay between reaction and phase behaviour in carbon dioxide hydrogenation to methanol. *ChemSusChem* **10**, 1166–1174 (2017).

71. Kondratenko, E. V., Mul, G., Baltrusaitis, J., Larrazabal, G. O. & Perez-Ramirez, J. Status and perspectives of CO$_2$ conversion into fuels and chemicals by catalytic, photocatalytic and electrocatalytic processes. *Energy Environ. Sci.* **6**, 3112–3135 (2013).

72. Olah, G. A. Beyond oil and gas: the methanol economy. *Angew. Chem. Int. Ed.* **44**, 2636–2639 (2005).

73. Lal, R. Soil carbon sequestration to mitigate climate change. *Geoderma* **123**, 1–22 (2004).

74. Lal, R., Negassa, W. & Lorenz, K. Carbon sequestration in soil. *Curr. Opin. Environ. Sustain.* **15**, 79–86 (2015).

75. Williamson, P. Emissions reduction: scrutinize CO$_2$ removal methods. *Nature* **530**, 153–155 (2016).

76. Marshall, C. In Switzerland, a giant new machine is sucking carbon directly from the air. *Science* https://doi.org/10.1126/science.aan6915 (2017).

77. Man, I. C. et al. Universality in oxygen evolution electrocatalysis on oxide surfaces. *ChemCatChem* **3**, 1159–1165 (2011).

78. Montoya, J. H., Tsai, C., Vojvodic, A. & Nørskov, J. K. The challenge of electrochemical ammonia synthesis: a new perspective on the role of nitrogen scaling relations. *ChemSusChem* **8**, 2180–2186 (2015).

79. Studt, F. et al. Discovery of a Ni–Ga catalyst for carbon dioxide reduction to methanol. *Nat. Chem.* **6**, 320–324 (2014).

80. Benck, J. D., Hellstern, T. R., Kibsgaard, J., Chakthranont, P. & Jaramillo, T. F. Catalyzing the hydrogen evolution reaction (HER) with molybdenum sulfide nanomaterials. *ACS Catal.* **4**, 3957–3971 (2014).

81. Montoya, J. H. et al. Materials for solar fuels and chemicals. *Nat. Mater.* **16**, 70–81 (2017).

82. Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-learning-augmented chemisorption model for CO$_2$ electroreduction catalyst screening. *J. Phys. Chem. Lett.* **6**, 3528–3533 (2015).

83. Ulissi, Z. W., Medford, A. J., Bligaard, T. & Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.* **8**, 14621 (2017).

84. Montoya, J. H. & Persson, K. A. A high-throughput framework for determining adsorption energies on solid surfaces. *Comput. Mater.* **3**, 14 (2017).

85. Lysgaard, S., Landis, D. D., Bligaard, T. & Vegge, T. Genetic algorithm procreation operators for alloy nanoparticle catalysts. *Top. Catal.* **57**, 33–39 (2014).

86. Vilhelmsen, L. B. & Hammer, B. A genetic algorithm for first principles global structure optimization of supported nano structures. *J. Chem. Phys.* **141**, 044711 (2014).

87. Rosenbrock, C. W., Homer, E. R., Csányi, G. & Hart, G. L. W. Discovering the building blocks of atomic systems using machine learning: application to grain boundaries. *Comput. Mater.* **3**, 29 (2017).

88. Jinnouchi, R. & Asahi, R. Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning

algorithm. *J. Phys. Chem. Lett.* **8**, 4279–4283 (2017).

89. Greeley, J., Jaramillo, T. F., Bonde, J., Chorkendorff, I. & Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **5**, 909–913 (2006).

90. García-Mota, M., Vojvodic, A., Abild-Pedersen, F. & Nørskov, J. K. Electronic origin of the surface reactivity of transition-metal-doped TiO₂(110). *J. Phys. Chem. C* **117**, 460–465 (2013).

91. Hummelshøj, J. S., Abild-Pedersen, F., Studt, F., Bligaard, T. & Nørskov, J. K. CatApp: a web application for surface chemistry and heterogeneous catalysis. *Angew. Chem. Int. Ed.* **51**, 272–274 (2012).

92. Tran, R. et al. Surface energies of elemental crystals. *Sci. Data* **3**, 160080 (2016).

93. Kalidindi, S. R., Medford, A. J. & McDowell, D. L. Vision for data and informatics in the future materials innovation ecosystem. *JOM* **68**, 2126–2137 (2016).

94. Green, M. A. Commercial progress and challenges for photovoltaics. *Nat. Energy* **1**, 15015 (2016).

95. Haegel, N. M. et al. Terawatt-scale photovoltaics: trajectories and challenges. *Science* **356**, 141–143 (2017).

96. Kojima, A., Teshima, K., Shirai, Y. & Miyasaka, T. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. Am. Chem. Soc.* **131**, 6050–6051 (2009).

97. Tan, H. et al. Efficient and stable solution-processed planar perovskite solar cells via contact passivation. *Science* **355**, 722–726 (2017).

98. National Renewable Energy Laboratory. Best research-cell efficiencies. *National Renewable Energy Laboratory* www.nrel.gov/pv/assets/images/efficiency_chart.jpg (2016).

99. Shin, S. S. et al. Colloidally prepared La-doped BaSnO₃ electrodes for efficient, photostable perovskite solar cells. *Science* **356**, 167–171 (2017).

100. Li, G., Zhu, R. & Yang, Y. Polymer solar cells. *Nat. Photonics* **6**, 153–161 (2012).

101. Gaudiana, R. & Brabec, C. J. Fantastic plastic. *Nat. Photonics* **2**, 287 (2008).

102. Hoth, C. N., Schilinsky, P., Choulis, S. A., Balasubramanian, S. & Brabec, C. J. in *Applications of Organic and Printed Electronics* (ed. Cantatore, E.) 27–56 (Springer US, Boston, MA, 2013).

103. Al-Ibrahim, M., Roth, H.-K., Zhokhavets, U., Gobsch, G. & Sensfuss, S. Flexible large area polymer solar cells based on poly(3-hexylthiophene)/fullerene. *Sol. Energy Mater. Sol. Cells* **85**, 13–20 (2005).

104. Kaltenbrunner, M. et al. Ultrathin and lightweight organic solar cells with high flexibility. *Nat. Commun.* **3**, 770 (2012).

105. Schubert, M. B. & Werner, J. H. Flexible solar cells for clothing. *Mater. Today* **9**, 42–50 (2006).

106. Salvador, M. et al. Suppressing photooxidation of conjugated polymers and their blends with fullerenes through nickel chelates. *Energy Environ. Sci.* **10**, 2005–2016 (2017).

107. Henemann, A. BIPV: built-in solar energy. *Renew. Energy Focus* **9**, 14–19 (2008).

108. Azzopardi, B. et al. Economic assessment of solar electricity production from organic-based photovoltaic modules in a domestic environment. *Energy Environ. Sci.* **4**, 3741–3753 (2011).

109. Li, N. & Brabec, C. J. Washing away barriers. *Nat. Energy* **2**, 772–773 (2017).

110. Perea, J. D. et al. Introducing a new potential figure of merit for evaluating microstructure stability in photovoltaic polymer-fullerene blends. *J. Phys. Chem. C* **121**, 18153–18161 (2017).

111. Teichler, A. et al. Combinatorial screening of polymer:fullerene blends for organic solar cells by inkjet printing. *Adv. Energy Mater.* **1**, 105–114 (2011).

112. Chen, S. et al. Exploring the stability of novel wide bandgap perovskites by a robot based high throughput approach. *Adv. Energy Mater.* **8**, 1701543 (2018).

113. Lawrence Livermore National Laboratory. Energy flow charts. *LLNL Flow Charts* https://flowcharts.llnl.gov/ (2016).

114. Zebarjadi, M., Esfarjani, K., Dresselhaus, M. S., Ren, Z. F. & Chen, G. Perspectives on thermoelectrics: from fundamentals to device applications. *Energy Environ. Sci.* **5**, 5147–5162 (2012).

115. Biswas, K. et al. High-performance bulk thermoelectrics with all-scale hierarchical architectures. *Nature* **489**, 414–418 (2012).

116. Aydemir, U. et al. YCuTe₂: a member of a new class of thermoelectric materials with CuTe₂-based layered structure. *J. Mater. Chem. A* **4**, 2461–2472 (2016).

117. Chen, W. et al. Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment. *J. Mater. Chem. C* **4**, 4414–4426 (2016).

118. Jain, A., Hautier, G., Ong, S. P. & Persson, K. New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships. *J. Mater. Res.* **31**, 977994 (2016).

119. Pohls, J.-H. et al. Metal phosphides as potential thermoelectric materials. *J. Mater. Chem. C* **5**, 12441–12456 (2017).

120. Faghaninia, A. et al. A computational assessment of the electronic, thermoelectric, and defect properties of bournonite (CuPbSbS₃) and related substitutions. *Phys. Chem. Chem. Phys.* **19**, 6743–6756 (2017).

121. Kim, H. M., Shao, L., Zhang, K. & Pipe, K. P. Engineered doping of organic semiconductors for enhanced thermoelectric efficiency. *Nat. Mater.* **12**, 719–723 (2013).

122. Russ, B., Glaudell, A., Urban, J. J., Chabinyc, M. L. & Segalman, R. A. Organic thermoelectric materials for energy harvesting and temperature control. *Nat. Rev. Mater.* **1**, 16050 (2016).

123. Sun, L. et al. A microporous and naturally nanostructured thermoelectric metal–organic framework with ultralow thermal conductivity. *Joule* **1**, 168–177 (2017).

124. Ürge-Vorsatz, D., Cabeza, L. F., Serrano, S., Barreneche, C. & Petrichenko, K. Heating and cooling energy trends and drivers in buildings. *Renew. Sustain. Energy Rev.* **41**, 85–98 (2015).

125. Waqas, A. & Din, Z. U. Phase change material (PCM) storage for free cooling of buildings — a review. *Renew. Sustain. Energy Rev.* **18**, 607–625 (2013).

126. Memon, S. A. Phase change materials integrated in building walls: a state of the art review. *Renew. Sustain. Energy Rev.* **31**, 870–906 (2014).

127. Baetens, R., Jelle, B. P. & Gustavsen, A. Phase change materials for building applications: a state-of-the-art review. *Energy Build.* **42**, 1361–1368 (2010).

128. Koebel, M., Rigacci, A. & Achard, P. Aerogel-based thermal superinsulation: an overview. *J. Sol-Gel Sci. Technol.* **63**, 315–339 (2012).

129. Bendahou, D., Bendahou, A., Seantier, B., Grohens, Y. & Kaddami, H. Nano-fibrillated cellulose-zeolites based new hybrid composites aerogels with super thermal insulating properties. *Ind. Crops Prod.* **65**, 374–382 (2015).

130. Seantier, B., Bendahou, D., Bendahou, A., Grohens, Y. & Kaddami, H. Multi-scale cellulose based new bio-aerogel composites with thermal super-insulating and tunable mechanical properties. *Carbohydr. Polym.* **138**, 335–348 (2016).

131. Wicklein, B. et al. Thermally insulating and fire-retardant lightweight anisotropic foams based on nanocellulose and graphene oxide. *Nat. Nanotechnol.* **10**, 277–283 (2015).

132. Wang, Y., Runnerstrom, E. L. & Milliron, D. J. Switchable materials for smart windows. *Annu. Rev. Chem. Bio. Eng.* **7**, 283–304 (2016).

133. Runnerstrom, E. L., Llordes, A., Lounis, S. D. & Milliron, D. J. Nanostructured electrochromic smart windows: traditional materials and NIR-selective plasmonic nanocrystals. *Chem. Commun.* **50**, 10555–10572 (2014).

134. Kamalisarvestani, M., Saidur, R., Mekhilef, S. & Javadi, F. Performance, materials and coating technologies of thermochromic thin films on smart windows. *Renew. Sustain. Energy Rev.* **26**, 353–364 (2013).

135. Baetens, R., Jelle, B. P. & Gustavsen, A. Properties, requirements and possibilities of smart windows for dynamic daylight and solar energy control in buildings: a state-of-the-art review. *Sol. Energy Mater. Sol. Cells* **94**, 87–105 (2010).

136. DeForest, N. et al. United States energy and CO₂ savings potential from deployment of near-infrared electrochromic window glazings. *Build. Environ.* **89**, 107–117 (2015).

137. Monk, P. M. S. *The Viologens: Physicochemical Properties, Synthesis and Applications of the Salts of 4,4′-Bipyridine.* (Wiley, Weinheim, 1999).

138. Jasinski, R. J. n-Heptylviologen radical cation films on transparent oxide electrodes. *J. Electrochem. Soc.* **125**, 1619–1623 (1978).

139. Sammells, A. F. & Pujare, N. U. Electrochromic effects on heptylviologen incorporated within a solid polymer electrolyte cell. *J. Electrochem. Soc.* **133**, 1270–1271 (1986).

140. Akahoshi, H., Toshima, S. & Itaya, K. Electrochemical and spectroelectrochemical properties of polyviologen complex modified electrodes. *J. Phys. Chem.* **85**, 818–822 (1981).

141. Beaujuge, P. M. & Reynolds, J. R. Color control in π-conjugated organic polymers for use in electrochromic devices. *Chem. Rev.* **110**, 268–320 (2010).

142. Ribeiro, A. S. & Mortimer, R. J. Conjugated conducting polymers with electrochromic and fluorescent properties. *Electrochemistry* **13**, 21–49 (2016).

143. Kline, W. M., Lorenzini, R. G. & Sotzing, G. A. A review of organic electrochromic fabric devices. *Color. Technol.* **130**, 73–80 (2014).

144. Monk, P. M. S., Mortimer, R. J. & Rosseinsky, D. R. *Electrochromism: Fundamentals and Applications* (Wiley, Weinheim, 1995).

145. Mortimer, R. J. Electrochromic materials. *Ann. Rev. Mater. Res.* **41**, 241–268 (2011).

146. Xie, Y.-X., Zhao, W.-N., Li, G.-C., Liu, P.-F. & Han, L. A naphthalenediimide-based metal–organic framework and thin film exhibiting photochromic and electrochromic properties. *Inorg. Chem.* **55**, 549–551 (2016).

147. Wade, C. R., Li, M. & Dinca, M. Facile deposition of multicolored electrochromic metal–organic framework thin films. *Angew. Chem. Int. Ed.* **52**, 13377–13381 (2013).

148. AlKaabi, K., Wade, C. R. & Dincă, M. Transparent-to-dark electrochromic behavior in naphthalene-diimide-based mesoporous MOF-74 analogs. *Chem* **1**, 264–272 (2016).

149. Mjejri, I., Doherty, C. M., Rubio-Martinez, M., Drisko, G. L. & Rougier, A. Double-sided electrochromic device based on metal–organic frameworks. *ACS Appl. Mater. Interfaces* **9**, 39930–39934 (2017).

150. Mehlana, G. & Bourne, S. A. Unravelling chromism in metal–organic frameworks. *CrystEngComm* **19**, 4238–4259 (2017).

151. Gomez-Gualdron, D. A. et al. Computational design of metal–organic frameworks based on stable zirconium building units for storage and delivery of methane. *Chem. Mater.* **26**, 5632–5639 (2014).

152. Chung, Y. G. et al. In silico discovery of metal–organic frameworks for precombustion CO₂ capture using a genetic algorithm. *Sci. Adv.* **2**, e1600909 (2016).

153. Borboudakis, G. et al. Chemically intuited, large-scale screening of MOFs by machine learning techniques. *Comput. Mater.* **3**, 40 (2017).

154. Wilmer, C. E. et al. Large-scale screening of hypothetical metal–organic frameworks. *Nat. Chem.* **4**, 83–89 (2011).

155. Pardakhti, M., Moharreri, E., Wanik, D., Suib, S. L. & Srivastava, R. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). *ACS Comb. Sci.* **19**, 640–645 (2017).

156. Thackeray, M. M., Wolverton, C. & Isaacs, E. D. Electrical energy storage for transportation approaching the limits of, and going beyond, lithium-ion batteries. *Energy Environ. Sci.* **5**, 7854–7863 (2012).

157. Winsberg, J., Hagemann, T., Janoschka, T., Hager, M. D. & Schubert, U. S. Redox-flow batteries: from metals to organic redox-active materials. *Angew. Chem. Int. Ed.* **56**, 686–711 (2017).

158. González, A., Goikolea, E., Barrena, J. A. & Mysyk, R. Review on supercapacitors: technologies and materials. *Renew. Sustain. Energy Rev.* **58**, 1189–1206 (2016).

159. Goodenough, J. B. & Park, K. S. The Li-ion rechargeable battery: a perspective. *J. Am. Chem. Soc.* **135**, 1167–1176 (2013).

160. Choi, J. W. & Aurbach, D. Promise and reality of post-lithium-ion batteries with high energy densities. *Nat. Rev. Mater.* **1**, 16013 (2016).

161. Hautier, G., Fischer, C., Ehrlacher, V., Jain, A. & Ceder, G. Data mined ionic substitutions for the discovery of new compounds. *Inorg. Chem.* **50**, 656–663 (2011).

162. Hautier, G. et al. Phosphates as lithium-ion battery cathodes: an evaluation based on high-throughput ab initio calculations. *Chem. Mater.* **23**, 3495–3508 (2011).

163. Chen, H. et al. Carbonophosphates: a new family of cathode materials for Li-ion batteries identified computationally. *Chem. Mater.* **24**, 2009–2016 (2012).

164. Ermon, S., Xue, Y., Gomes, C. & Selman, B. Learning policies for battery usage optimization in electric vehicles. *Machine Learn.* **92**, 177–194 (2013).

165. Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M. & Dietmayer, K. Health diagnosis and remaining useful life prognosis of lithium-ion batteries using data-driven methods. *J. Power Sources* **239**, 680–688 (2013).

166. Waag, W., Fleischer, C. & Sauer, D. U. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. *J. Power Sources* **258**, 321–339 (2014).

167. Chaouachi, A., Kamel, R. M., Andoulsi, R. & Nagasaka, K. Multiobjective intelligent energy management for a microgrid. *IEEE Trans. Ind. Electron.* **60**, 1688–1699 (2013).

168. Huskinson, B. et al. A metal-free organic–inorganic aqueous flow battery. *Nature* **505**, 195–198 (2014).

169. Lin, K. et al. Alkaline quinone flow battery. *Science* **349**, 1529–1532 (2015).

170. Lin, K. et al. A redox-flow battery with an alloxazine-based organic electrolyte. *Nat. Energy* **1**, 16102 (2016).

171. Liu, T., Wei, X., Nie, Z., Sprenkle, V. & Wang, W. A total organic aqueous redox flow battery employing a low cost and sustainable methyl viologen anolyte and 4-HO-TEMPO catholyte. *Adv. Energy Mater.* **6**, 1501449 (2016).

172. Hu, B., DeBruler, C., Rhodes, Z. & Liu, T. L. Long-cycling aqueous organic redox flow battery (AORFB) toward sustainable and safe energy storage. *J. Am. Chem. Soc.* **139**, 1207–1214 (2017).

173. Beh, E. S. et al. A neutral pH aqueous organic–organometallic redox flow battery with extremely high capacity retention. *ACS Energy Lett.* **2**, 639–644 (2017).

174. Pyzer-Knapp, E. O., Suh, C., Gómez-Bombarelli, R., Aguilera-Iparraguirre, J. & Aspuru-Guzik, A. What is high-throughput virtual screening? A perspective from organic materials discovery. *Annu. Rev. Mater. Res.* **45**, 195–216 (2015).

175. Jain, A., Shin, Y. & Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **1**, 15004 (2016).

176. Hachmann, J. et al. The Harvard Clean Energy Project: large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2**, 2241–2251 (2011).

177. Hachmann, J. et al. Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry — the Harvard Clean Energy Project. *Energy Environ. Sci.* **7**, 698–704 (2014).

178. Er, S., Suh, C., Marshak, M. P. & Aspuru-Guzik, A. Computational design of molecules for an all-quinone redox flow battery. *Chem. Sci.* **6**, 885–893 (2015).

179. Gómez-Bombarelli, R. et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **15**, 1120–1127 (2016).

180. Sokolov, A. N. et al. From computational discovery to experimental characterization of a high hole mobility organic crystal. *Nat. Commun.* **2**, 437 (2011).

181. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).

182. Jain, A. et al. The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

183. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* **111**, 218–230 (2016).

184. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).

185. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).

186. Curtarolo, S. et al. AFLOWLIB. ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).

187. Hautier, G. et al. Novel mixed polyanions lithium-ion battery cathode materials predicted by high-throughput ab initio computations. *J. Mater. Chem.* **21**, 17147–17153 (2011).

188. Kirklin, S., Chan, M. K. Y., Trahey, L., Thackeray, M. M. & Wolverton, C. High-throughput screening of high-capacity electrodes for hybrid Li-ion–Li–O_2 cells. *Phys. Chem. Chem. Phys.* **16**, 22073–22082 (2014).

189. Qu, X. et al. The Electrolyte Genome Project: a big data approach in battery materials discovery. *Comput. Mater. Sci.* **103**, 56–67 (2015).

190. Aykol, M. et al. High-throughput computational design of cathode coatings for Li-ion batteries. *Nat. Commun.* **7**, 13779 (2016).

191. Toher, C. et al. High-throughput computational screening of thermal conductivity, Debye temperature, and Gruneisen parameter using a quasiharmonic Debye model. *Phys. Rev. B* **90**, 174107 (2014).

192. Wu, Y., Lazic, P., Hautier, G., Persson, K. & Ceder, G. First principles high throughput screening of oxynitrides for water-splitting photocatalysts. *Energy Environ. Sci.* **6**, 157–168 (2013).

193. Khatami, S. N. & Aksamija, Z. Lattice thermal conductivity of the binary and ternary group-IV alloys Si-Sn, Ge-Sn, and Si-Ge-Sn. *Phys. Rev. Appl* **6**, 014015 (2016).

194. Compton, W. D. & Schulman, J. H. *Color Centers in Solids* 2 (Pergamon, Oxford, 1962).

195. Ding, H. et al. Computational approach for epitaxial polymorph stabilization through substrate selection. *ACS Appl. Mater. Interfaces* **8**, 13086–13093 (2016).

196. Dunstan, M. T. et al. Large scale computational screening and experimental discovery of novel materials for high temperature CO_2 capture. *Energy Environ. Sci.* **9**, 1346–1360 (2016).

197. Zhu, H. et al. Computational and experimental investigation of $TmAgTe_2$ and XYZ_2 compounds, a new group of thermoelectric materials identified by first-principles high-throughput screening. *J. Mater. Chem. C* **3**, 10554–10565 (2015).

198. Pyzer-Knapp, E. O., Li, K. & Aspuru-Guzik, A. Learning from the Harvard Clean Energy Project: the use of neural networks to accelerate materials discovery. *Adv. Funct. Mater.* **25**, 6495–6502 (2015).

199. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big data of materials science: critical role of the descriptor. *Phys. Rev. Lett.* **114**, 105503 (2015).

200. Segler, M. H. S., Kogej, T., Tyrchan, C. & Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **4**, 120–131 (2018).

201. Ikebata, H., Hongo, K., Isomura, T., Maezono, R. & Yoshida, R. Bayesian molecular design with a chemical language model. *J. Comput. Aided Mol. Des.* **31**, 379–391 (2017).

202. Kadurin, A. et al. The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* **8**, 10883–10890 (2017).

203. Podlewska, S., Czarnecki, W. M., Kafel, R. & Bojarski, A. J. Creating the new from the old: combinatorial libraries generation with machine-learning-based compound structure optimization. *J. Chem. Inf. Model.* **57**, 133–147 (2017).

204. Tibbetts, K. M., Feng, X.-J. & Rabitz, H. Exploring experimental fitness landscapes for chemical synthesis and property optimization. *Phys. Chem. Chem. Phys.* **19**, 4266–4287 (2017).

205. Moore, K. W. et al. Universal characteristics of chemical synthesis and property optimization. *Chem. Sci.* **2**, 417–424 (2011).

206. Moore, K. W. et al. Why is chemical synthesis and property optimization easier than expected? *Phys. Chem. Chem. Phys.* **13**, 10048–10070 (2011).

207. Ping Ong, S., Wang, L., Kang, B. & Ceder, G. Li–Fe–P–O_2 phase diagram from first principles calculations. *Chem. Mater.* **20**, 1798–1807 (2008).

208. Langer, J. S. Models of pattern formation in first-order phase transitions. *Dir. Condens. Matt. Phys.* **1**, 165–186 (1986).

209. Lee, D. D., Choy, J. H. & Lee, J. K. Computer generation of binary and ternary phase diagrams via a convex hull method. *J. Phase Equilib.* **13**, 365–372 (1992).

210. Pourbaix, M. *Atlas of Electrochemical Equilibria in Aqueous Solutions* 1 (Pergamon, Oxford, 1966).

211. Dannatt, C. W. & Ellingham, H. J. T. Roasting and reduction processes. Roasting and reduction processes-a general survey. *Discuss. Faraday Soc* **4**, 126–139 (1948).

212. Spencer, P. A brief history of CALPHAD. *Calphad* **32**, 1–8 (2008).

213. Phillips, R. *Crystals, Defects and Microstructures: Modeling Across Scales* (Cambridge Univ. Press, Cambridge, 2001).

214. Goyal, A., Gorai, P., Peng, H., Lany, S. & Stevanovic, V. A computational framework for automation of point defect calculations. Preprint at *arXiv*, 1611.00825 (2016).

215. Gomberg, J. A., Medford, A. J. & Kalidindi, S. R. Extracting knowledge from molecular mechanics simulations of grain boundaries using machine learning. *Acta Mater.* **133**, 100–108 (2017).

216. El-Awady, J. A. Unravelling the physics of size-dependent dislocation-mediated plasticity. *Nat. Commun.* **6**, 5926 (2015).

217. Wu, H., Mayeshiba, T. & Morgan, D. Dataset for high-throughput ab-initio dilute solute diffusion database. *Globus* https://doi.org/10.18126/M2X59R (2016).

218. Toher, C. et al. Combining the AFLOW GIBBS and elastic libraries to efficiently and robustly screen thermomechanical properties of solids. *Phys. Rev. Mater.* **1**, 015401 (2017).

219. de Jong, M. et al. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, 150009 (2015).

220. Sun, W. et al. The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225 (2016).

221. Bartók, A. P. et al. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **3**, e1701816 (2017).

222. Segler, M. H. S., Preuss, M. & Waller, M. P. Learning to plan chemical syntheses. Preprint at *arXiv*, 1708.04202 (2017).

223. Corey, E. J. & Jorgensen, W. L. Computer-assisted synthetic analysis. Synthetic strategies based on appendages and the use of reconnective transforms. *J. Am. Chem. Soc.* **98**, 189–203 (1976).

224. Corey, E. J. & Wipke, W. T. Computer-assisted design of complex organic syntheses. *Science* **166**, 178–192 (1969).

225. Pensak, D. A. & Corey, E. J. in *Computer-Assisted Organic Synthesis* (eds Wipke, W. T. & Howe, W. J.) 1–32 (American Chemical Society, Washington, DC, 1977).

226. Wipke, W. T. & Howe, W. J. *Computer-Assisted Organic Synthesis* (American Chemical Society, Washington, DC, 1977).

227. Jorgensen, W. L. et al. CAMEO: a program for the logical prediction of the products of organic reactions. *Pure Appl. Chem.* **62**, 1921–1932 (1990).

228. Gasteiger, J. & Jochum, C. EROS a computer program for generating sequences of reactions. *Organic Compunds* **74**, 93–126 (1978).

229. Satoh, H. & Funatsu, K. SOPHIA, a knowledge base-guided reaction prediction system — utilization of a knowledge base derived from a reaction database. *J. Chem. Inf. Comp. Sci.* **35**, 34–44 (1995).

230. Gelernter, H. L. et al. Empirical explorations of SYNCHEM. *Science* **197**, 1041–1049 (1977).

231. Pence, H. E. & Williams, A. ChemSpider: an online chemical information resource. *J. Chem. Ed.* **87**, 1123–1124 (2010).

232. Akhondi, S. A. et al. Annotated chemical patent corpus: a gold standard for text mining. *PLOS One* **9**, e107477 (2014).

233. Bøgevig, A. et al. Route design in the 21st century: the ICSYNTH software tool as an idea generator for synthesis prediction. *Org. Process Res. Dev.* **19**, 357–368 (2015).

234. Szymkuć, S. et al. Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).

235. Bergeler, M., Simm, G. N., Proppe, J. & Reiher, M. Heuristics-guided exploration of reaction mechanisms. *J. Chem. Theory Comput.* **11**, 5712–5722 (2015).

236. Kayala, M. A. & Baldi, P. ReactionPredictor: prediction of complex chemical reactions at the mechanistic level using machine learning. *J. Chem. Inf. Model.* **52**, 2526–2540 (2012).

237. Wei, J. N., Duvenaud, D. & Aspuru-Guzik, A. Neural networks for the prediction of organic chemistry reactions. *ACS Cent. Sci.* **2**, 725–732 (2016).

238. Segler, M. H. S. & Waller, M. P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chem. Eur. J.* **23**, 5966–5971 (2017).

239. Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H. & Jensen, K. F. Prediction of organic reaction outcomes using machine learning. *ACS Cent. Sci.* **3**, 434–443 (2017).

240. Duvenaud, D. K. et al. in *Advances in Neural Information Processing Systems* (eds Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 2224–2232 (Curran Associates, 2015).

241. Peplow, M. Organic synthesis: the robo-chemist. *Nature* **512**, 20–22 (2014).

242. Nicolaou, C. A., Watson, I. A., Hu, H. & Wang, J. The Proximal Lilly Collection: mapping, exploring and exploiting feasible chemical space. *J. Chem. Inf. Model.* **56**, 1253–1266 (2016).

243. Godfrey, A. G., Masquelin, T. & Hemmerle, H. A remote-controlled adaptive Medchem Lab: an innovative approach to enable drug discovery in the 21*st* century. *Drug Discov. Today* **18**, 795–802 (2013).

244. Nicolaou, K. C., Hanko, R. & Hartwig, W. in *Handbook of Combinatorial Chemistry* (eds Nicolaou, K. C., Hanko, R. & Hartwig, W.) 1–9 (Wiley-VCH, Weinheim, 2005).

245. Shevlin, M. Practical high-throughput experimentation for chemists. *ACS Med. Chem. Lett.* **8**, 601–607 (2017).

246. Weber, A., von Roedern, E. & Stilz, H. U. SynCar: an approach to automated synthesis. *J. Comb. Chem.* **7**, 178–184 (2005).

247. Prabhu, G. R. D. & Urban, P. L. The dawn of unmanned analytical laboratories. *Trends Anal. Chem.* **88**, 41–52 (2017).

248. Ley, S. V., Fitzpatrick, D. E., Myers, R. M., Battilocchio, C. & Ingham, R. J. Machine-assisted organic synthesis. *Angew. Chem. Int. Ed.* **54**, 10122–10136 (2015).

249. Pastre, J. C., Browne, D. L. & Ley, S. V. Flow chemistry syntheses of natural products. *Chem. Soc. Rev.* **42**, 8849–8869 (2013).

250. Adamo, A. et al. On-demand continuous-flow production of pharmaceuticals in a compact, reconfigurable system. *Science* **352**, 61–67 (2016).

251. Rasheed, M. & Wirth, T. Intelligent microflow: development of self-optimizing reaction systems. *Angew. Chem. Int. Ed.* **50**, 357–358 (2011).

252. Buitrago Santanilla, A. et al. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49–53 (2015).

253. Nelson, J. D. in *Practical Synthetic Organic Chemistry* (ed. Caron, S.) 1–71 (John Wiley & Sons, Hoboken, 2011).

254. Vaidyanathan, R. & Wager, C. B. in *Practical Synthetic Organic Chemistry* (ed. Caron, S.) 73–165 (John Wiley & Sons, Hoboken, 2011).

255. Caron, S. et al. in *Practical Synthetic Organic Chemistry* (ed. Caron, S.) 279–340 (John Wiley & Sons, Hoboken, 2011).

256. Ripin, D. H. B. in *Practical Synthetic Organic Chemistry* (ed. Caron, S.) 341–381; 493–556 (John Wiley & Sons, Hoboken, 2011).

257. Pouliot, J.-R., Grenier, F., Blaskovits, J. T., Beauprè, S. & Leclerc, M. Direct (hetero)arylation polymerization: simplicity for conjugated polymer synthesis. *Chem. Rev.* **116**, 14225–14274 (2016).

258. Woerly, E. M., Roy, J. & Burke, M. D. Synthesis of most polyene natural product motifs using just 12 building blocks and one coupling reaction. *Nat. Chem.* **6**, 484–491 (2014).

259. Service, R. F. The synthesis machine. *Science* **347**, 1190–1193 (2015).

260. Li, J. et al. Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).

261. Maiwald, M., Fischer, H. H., Kim, Y.-K., Albert, K. & Hasse, H. Quantitative high-resolution on-line NMR spectroscopy in reaction and process monitoring. *J. Magn. Reson.* **166**, 135–146 (2004).

262. Baranczak, A. et al. Integrated platform for expedited synthesis–purification–testing of small molecule libraries. *ACS Med. Chem. Lett.* **8**, 461–465 (2017).

263. Green, M. L. et al. Fulfilling the promise of the materials genome initiative with high- throughput experimental methodologies. *Appl. Phys. Rev.* **4**, 011105 (2017).

264. Xiang, X. D. et al. A combinatorial approach to materials discovery. *Science* **268**, 1738–1740 (1995).

265. Tsui, F. & Ryan, P. Combinatorial molecular beam epitaxy synthesis and char- acterization of magnetic alloys. *Appl. Surf. Sci.* **189**, 333–338 (2002).

266. Wang, Q., Itaka, K., Minami, H., Kawaji, H. & Koinuma, H. Combinatorial pulsed laser deposition and thermoelectricity of (La$_{1-x}$Ca$_x$)VO$_3$ composition-spread films. *Sci. Technol. Adv. Mater.* **5**, 543–547 (2004).

267. Chang, K.-S., Aronova, M. & Takeuchi, I. Combinatorial pulsed laser deposition using a compact high-throughout thin-film deposition flange. *Appl. Surf. Sci.* **223**, 224–228 (2004).

268. Takeuchi, I. in *Pulsed Laser Deposition of Thin Films* (ed. Eason, R.) 161–176 (John Wiley & Sons, Hoboken, 2006).

269. Kim, D. H. et al. Combinatorial pulsed laser deposition of Fe, Cr, Mn, and Ni-substituted SrTiO$_3$ films on Si substrates. *ACS Comb. Sci.* **14**, 179–190 (2012).

270. Havelia, S. et al. Combinatorial substrate epitaxy: a new approach to growth of complex metastable compounds. *CrystEngComm* **15**, 5434–5441 (2013).

271. Sun, X. Y. et al. Combinatorial pulsed laser deposition of magnetic and magneto-optical Sr(Ga$_x$Ti$_y$Fe$_{0.34-0.40}$)O$_{3-\delta}$ perovskite films. *ACS Comb. Sci.* **16**, 640–646 (2014).

272. Kadhim, A. et al. Development of combinatorial pulsed laser deposition for expedited device optimization in CdTe/CdS thin-film solar cells. *Int. J. Opt.* **2016**, 1696848 (2016).

273. Keifer, P. A. High-resolution NMR techniques for solid-phase synthesis and combinatorial chemistry. *Drug Discov. Today* **2**, 468–478 (1997).

274. Hamper, B. C. et al. High-throughput ^{1}H NMR and HPLC characterization of a 96-member substituted methylene malonamic acid library. *J. Comb. Chem.* **1**, 140–150 (1999).

275. Carter, C. F. et al. ReactIR Flow Cell: a new analytical tool for continuous flow chemical processing. *Org. Process Res. Dev.* **14**, 393–404 (2010).

276. Huang, H., Yu, H., Xu, H. & Ying, Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *J. Food. Eng.* **87**, 303–313 (2008).

277. Otani, M. et al. A high-throughput thermoelectric power-factor screening tool for rapid construction of thermoelectric property diagrams. *Appl. Phys. Lett.* **91**, 132102 (2007).

278. Kuo, T.-C., Malvadkar, N. A., Drumright, R., Cesaretti, R. & Bishop, M. T. High- throughput industrial coatings research at The Dow Chemical Company. *ACS Comb. Sci* **18**, 507–526 (2016).

279. Hepp, J., Machui, F., Egelhaaf, H.-J., Brabec, C. J. & Vetter, A. Automatized analysis of IR-images of photovoltaic modules and its use for quality control of solar cells. *Energy Sci. Eng.* **4**, 363–371 (2016).

280. Alstrup, J., Jørgensen, M., Medford, A. J. & Krebs, F. C. Ultra fast and parsimonious materials screening for polymer solar cells using differentially pumped slot-die coating. *ACS Appl. Mater. Interfaces* **2**, 2819–2827 (2010).

281. Guldal, N. S. et al. Real-time evaluation of thin film drying kinetics using an advanced, multi-probe optical setup. *J. Mater. Chem. C* **4**, 2178–2186 (2016).

282. Dragone, V., Sans, V., Henson, A. B., Granda, J. M. & Cronin, L. An autonomous organic reaction search engine for chemical reactivity. *Nat. Commun.* **8**, 15733 (2017).

283. Kitson, P. J. et al. Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals. *Science* **359**, 314–319 (2018).

284. Gutierrez, J. P. M., Hinkley, T., Taylor, J. W., Yanev, K. & Cronin, L. Evolution of oil droplets on a chemorobtic platform. *Nat. Commun.* **5**, 5571 (2014).

285. Krein, M., Huang, T. W., Morkowchuk, L., Agrafiotis, D. K. & Breneman, C. M. in *Statistical Modelling of Molecular Descriptors in QSAR/QSPR* (eds Dehmer, M., Varmuza, K., Bonchev, D. & Emmert-Streib, F.) 33–64 (Wiley-Blackwell, Weinheim, 2012).

286. Seffers, G. I. Scientists pick AI for lab partner. *AFCEA* https://www.afcea.org/content/scientists-pick-ai-lab-partner (2017).

287. Kaur, N. & Sood, S. K. An energy-efficient architecture for the Internet of Things (IoT). *IEEE Syst. J.* **11**, 796–805 (2017).

288. Jacoby, M. The future of low-cost solar cells. *Chem. Eng. News* **94**, 30–35 (2016).

289. Snyder, G. J. & Toberer, E. S. Complex thermoelectric materials. *Nat. Mater.* **7**, 105–114 (2008).

290. Korgel, B. A. Materials science: composite for smarter windows. *Nature* **500**, 278–279 (2013).

291. Mathews, C. Battery storage: power of good can flow in SA. *Financial Mail* https://www.businesslive.co.za/fm/fm-fox/2017-06-29-battery-storage-power-of-good-can-flow-in-sa/ (2017).

292. Zhang, C. et al. Thienobenzene-fused perylene bisimide as a non-fullerene acceptor for organic solar cells with a high open-circuit voltage and power conversion efficiency. *Mater. Chem. Front.* **1**, 749–756 (2017).

293. Yan, Y. G., Martin, J., Wong-Ng, W., Green, M. & Tang, X. F. A temperature dependent screening tool for high throughput thermoelectric characterization of combinatorial films. *Rev. Sci. Instrum.* **84**, 115110 (2013).

RELATED LINKS

Awesome Pipeline: https://github.com/pditommaso/awesome-pipeline

ChemPlanner: http://www.chemplanner.com/

Clean Energy Materials Innovation Challenge: http://mission-innovation.net/our- work/innovation-challenges/clean-energy-materials-challenge/

Climeworks: http://www.climeworks.com/

Dial-a-Molecule: http://generic.wordpress.soton.ac.uk/dial-a-molecule/

InfoChem: http://www.infochem.de/products/databases/spresi.shtml

Inorganic Crystal Structure Database: http://www2.fiz-karlsruhe.de/icsd_home.html

NIST high-throughput screening tool: https://www.nist.gov/laboratories/tools-instruments/high-throughput-combinatorial-screening-tool-characterization-thin

RDKit: Open-source cheminformatics http://www.rdkit.org

Reaxys: https://www.reaxys.com/

UniEnergy Technologies: http://www.uetechnologies.com/