

Design of functional and sustainable polymers assisted by artificial intelligence

Huan Tran^{1,2}, Rishi Gurnani², Chiho Kim^{1,2}, Ghanshyam Pilania³, Ha-Kyung Kwon⁴, Ryan P. Lively⁵ & Rampi Ramprasad^{1,2}✉

Abstract

Artificial intelligence (AI)-based methods continue to make inroads into accelerated materials design and development. Here, we review AI-enabled advances made in the subfield of polymer informatics, with a particular focus on the design of application-specific practical polymeric materials. We consider exemplar design attempts within a few critical and emerging application spaces, including materials designs for storing, producing and conserving energy, and those that can prepare us for a sustainable economy powered by recyclable and/or biodegradable polymers. AI-powered workflows help to efficiently search the staggeringly large chemical and configurational space of materials, using modern machine-learning (ML) algorithms to solve ‘forward’ and ‘inverse’ materials design problems. A theme explored throughout this Review is a practical informatics-based design protocol that involves creating a set of application-specific target property criteria, building ML model predictors for those relevant target properties, enumerating or generating a tangible population of viable polymers, and selecting candidates that meet design recommendations. The protocol is demonstrated for several energy- and sustainability-related applications. Finally, we offer our outlook on the lingering obstacles that must be overcome to achieve widespread adoption of informatics-driven protocols in industrial-scale materials development.

Sections

Introduction

Application-specific design of polymers

Critical next steps

Outlook

¹School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ²Matmerize Inc., Atlanta, GA, USA. ³GE Aerospace Research, Niskayuna, NY, USA. ⁴Toyota Research Institute, Los Altos, CA, USA. ⁵School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: rampi.ramprasad@mse.gatech.edu

Introduction

Polymeric materials have an ancient history, but the foundations of modern polymer science can be traced back to work by Hermann Staudinger in the 1920s. His groundbreaking idea¹ was that high-molecular-weight materials, such as rubber, cellulose and proteins, consist of lengthy chains formed by repeating molecular-size units linked by covalent bonds^{2,3}. This fundamental concept of modern polymer science played a crucial role in numerous remarkable discoveries and advancements (Fig. 1a), including the creation of innovative polymers such as polypropylene, neoprene, nylon, Teflon and Kevlar. These polymers, from daily packaging materials to high-tech device components, have permeated every aspect of our world^{4–13}. The ability to control essential parameters such as chemical structure, processing conditions and additives has enabled the development of synthetic polymers with diverse properties, ranging from rigidity to elasticity, a broad range of electrical conductivity, and permeability and selectivity to specific gases. This versatility stems from factors such as the structural diversity of organic materials, the exceptional synthetic ingenuity of chemists, and the vast chemical space that polymers occupy. Staudinger's seminal work, honoured with a Nobel Prize in 1953, laid the foundation for successive Nobel Prizes in this field.

Although many noteworthy polymeric materials have been discovered, developed and commercially deployed over the past century, the transition from concept to deployment has required years to decades even in the most successful instances (Fig. 1b). Several factors have contributed to this prolonged timeline. First, key concepts originate from the intuition and experience of a select few expert scientists and engineers. Pursuing these original ideas, either using physical experimentation or computer simulations, demands specialized skills, funds, resources and time; methodically exploring the vast chemical and/or processing space is non-trivial even within a restricted class of materials. The new material must satisfy various success metrics, encompassing properties, performance, cost, safety and supply chain considerations. And finally, attaining a satisfactory end point in a timely manner, ahead of the competition, necessitates prioritizing options such as the fastest or most cost-effective synthetic pathways. These considerations lead to substantial trial-and-error activities, missed opportunities and a sizeable reliance on serendipity.

It is tempting to imagine a future in which materials intuition, experience, and the vast repository of data and knowledge can be encoded and embedded in a powerful artificial intelligence (AI) expert system. This could not only safeguard against the loss (or neglect) of valuable assets, be they data or knowledge, but also hold the promise of continuous improvement, rapid and reliable property predictions, informed decision-making and democratization – making expertise readily accessible to anyone at any time. This philosophy has catalysed the emergence of several materials informatics ecosystems, summarized in Table 1, around the globe in the past decade or so¹⁴. Fuelled in part by the [Materials Genome Initiative](#), these ecosystems serve to complement, augment and elevate the impact of empirical or computation-based materials research. Various indicators suggest that this vision is gaining traction in industry^{15,16}, driven by the perception and expectation that such AI-based knowledge systems can substantially reduce both the number and timelines of iterative cycles preceding the deployment of new materials.

This Review centres on polymer informatics^{17–22}, specifically delving into AI-driven polymer designs tailored for various applications. The roots of polymer informatics can be traced back decades, initially emerging as 'group contribution' methods²³. These methods used

numerical representations of polymers based on their molecular fragments, mapped onto properties through linear regression for rapid predictions. Today's polymer informatics ecosystems have evolved considerably. First, contemporary polymer property datasets are larger, more reliable and cover many more properties than the group contribution methods. Second, recent numerical representation (or fingerprinting) schemes^{24–27} are much more comprehensive, in some cases^{24,25} incorporating thousands of descriptors across multiple length scales, ranging from the atomic to block to chain levels. They are also scalable (to handle search spaces of over a billion candidates), leveraging modern transformer-based^{28,29}, graph-based^{30–32} and chemical language models³³. Finally, machine-learning (ML) algorithms that map polymer fingerprints to properties (leading to predictive models) can handle extensive datasets, are generalizable and interpretable, and accommodate data from diverse sources and fidelity levels^{24,25,34}.

These rapid and reliable property prediction models are critical for materials informatics. However, materials design requires more than prediction; it must involve the ability to 'invert' the prediction process and recommend materials that align with target properties or performances^{35–37}. Over the past decade, various inverse methods, including high-throughput screening^{38,39}, Monte Carlo schemes⁴⁰, recommender systems⁴¹, Bayesian optimization^{42,43}, particle swarm optimization^{44–46}, evolutionary or genetic algorithms^{47,48}, syntax-directed variational autoencoders^{49,50} and graph-to-graph translation³⁰, have proven effective at proposing candidate materials, particularly for polymers. However, a key challenge is ensuring that these materials are synthetically feasible. To tackle this, an emerging approach, which we call 'virtual forward synthesis' (VFS), could leverage millions of commercially available or easy-to-synthesize monomers and insert them in several hundreds of known polymerization reaction templates to digitally generate any number of synthetically accessible polymers^{51–53}.

A robust strategy for use-inspired, application-driven, synthetically accessible polymer design involves several steps (Box 1). It begins with defining a set of screening criteria based on desired property values. A vast search space is then defined, potentially involving a large library of polymers produced through VFS or similar methods. In parallel, predictive ML models for these key properties are developed using sufficiently large and diverse datasets of measured or computed values⁵⁴. The models predict key properties for polymers in the search space, and those meeting desired values are selected as potential candidates. The design loop is 'closed' by testing the recommended candidates via physical experiments. Success is declared if the candidates meet the required property or performance criteria. If not, the design loop cycle repeats.

In this Review, we highlight several polymer design endeavours undertaken in the past few years that have benefited from applying AI or informatics methodologies. These case studies span polymer dielectrics for energy storage^{7,55,56}, fuel-cell membranes and ionomers⁵⁷, solid polymer electrolytes for batteries^{58–61}, gas^{10,62} and liquid mixture^{63–66} separation membranes, biodegradable polymers⁶⁷ and depolymerizable polymers^{68–71}. We then examine challenges and opportunities on the horizon, touching on complex scenarios involving composites, formulations and cross-linked polymers; the autonomous extraction of property data from the exponentially increasing literature using natural language processing techniques; and the use of computational–experimental information fusion and multifidelity methods to produce and leverage data covering ever-greater chemical spaces. Ultimately, transitioning from successful laboratory-scale synthesis to commercialization is the definitive validation of the real-world value of polymer informatics.

a Selected milestones of polymer science

- **1833:** Berzelius coins the term 'polymer'
- **1844:** Goodyear patents the vulcanization of rubber
- **1846:** Schonbein synthesizes nitrocellulose
- **1856:** Parkes patents the first thermoplastic
- **1892:** Cross, Bevan and Beadle patent 'viscose'
- **1907:** Baekeland creates the first synthetic plastic
- **1912:** Klatte invents the production process of PVC
- **1920:** Staudinger publishes 'uber polymerization'
- **1933:** Fawcett and Gibson discover LDPE
- **1937:** Bayer discovers polyurethane
- **1938:** Carothers patents nylon, i.e., Polyamide 6.6, and Plunkett discovers Teflonane
- **1944:** McIntire discovers polystyrene
- **1954:** Ziegler–Natta catalyst discovered for polypropylene production
- **1965:** Kwolek creates Kevlar
- **1977:** Heeger, MacDiarmid and Shirakawa develop polyacetylene, the first conductive polymer
- **2011:** Launch of Materials Genome Initiative
- **2020:** 100-year anniversary of modern polymer science

b Typical deployment progress of some notable polymers

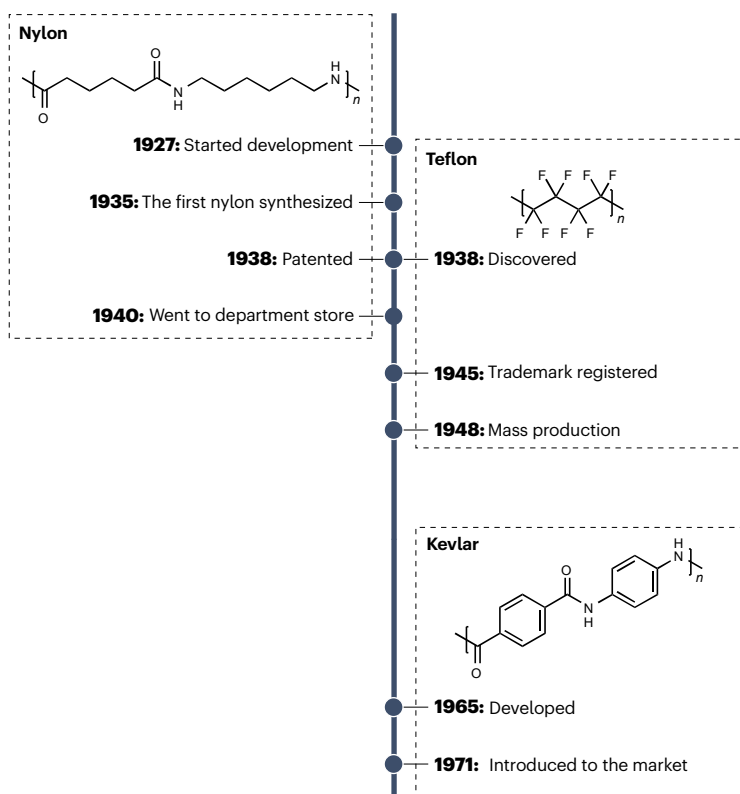


Fig. 1 | Polymer innovations over the past two centuries. **a**, Selected chronological milestones in polymer science. **b**, Traditional transition from concept to deployment of some notable polymeric materials. LDPE, low-density polyethylene; PVC, poly(vinyl chloride).

Application-specific design of polymers

Dielectric polymers for electrostatic energy storage

The escalating global demands for electric power and energy storage present profound challenges, one of the foremost being the design of materials for use in electrostatic capacitor devices that can withstand extreme electric fields and temperatures^{7,72–79}. The selection of suitable dielectric materials, temperatures and electric fields for such applications is bound by the limitations imposed by electrical breakdown. Beyond a critical field intensity, accelerating electron cascades within an insulator lead to electrical discharge and system failure. This behaviour is exacerbated at elevated temperatures and is inherently governed by the material's chemical composition and morphology⁸⁰.

Polymeric dielectric materials, owing to their graceful failure modes, are the materials of choice for capacitive energy storage across transportation, aerospace, energy and defence sectors. In comparison to more extensively discussed energy-storage devices such as batteries, fuel cells and supercapacitors, electrostatic capacitors present unparalleled power density^{7,55,72–76} (Fig. 2a). This positions electrostatic capacitors as particularly advantageous for a wide array of applications,

including hybrid and all-electric systems, pulsed power systems, wind pitch control, aircraft launchers and space exploration^{7,72–76}.

Current high-power capacitors use biaxially oriented polypropylene (BOPP) as the dielectric, a material that has been used for over three decades. Its long-term presence is due to its high electrical breakdown strength of over 700 MV m^{−1} at room temperature, low cost for mass production, and considerable investment from the academic community, industry and supply chain. BOPP, along with similar commercial polyolefins, also exhibits low dielectric loss and a substantial electronic bandgap, attributed, in part, to the absence of π -stacking moieties. Despite these features, these materials have a low dielectric constant, resulting in a diminished electrostatic energy density – a critical 'figure of merit' for this application. At room temperature, BOPP registers a baseline energy density of 5 J cm^{−3}, which rapidly degrades with increasing temperature.

Rational materials design approaches to surpass BOPP in electrostatic energy storage have proven successful^{55,56,81–83}, driven by the establishment of clear property-based screening criteria and the integration of computational methodologies that primarily use density

Table 1 | Notable polymer informatics ecosystems

Name	Category	Short description	Location	Refs.
PolyInfo	Database	~32k homopolymers, copolymers, polymer blends and composites, ~500k experimental data points	https://polymer.nims.go.jp	269
PI1M	Database, open	~1M polymers generated from generative models trained on ~12k actual polymers	https://github.com/RUIMINMA1996/PI1M	270
CHEMnetBASE	Database, open	Polymers and properties	https://poly.chemnetbase.com	
polyVERSE	Database, open	~200M generated polymers, ~10k properties computed for synthesized polymers	https://github.com/Ramprasad-Group/polyVERSE	33,54,80,271
Polymer Scholar	Database, open	~300k polymer property records extracted from literature	https://polymerscholar.org	192,243,244
OMG	Database, open	Open Macromolecular Genome, ~12M linear homopolymers created by commercially available monomers and 17 canonical polymerizations	https://zenodo.org/records/7556992	51
HTPMD	Database, open	6,286 high-throughput MD trajectories of amorphous polymer electrolytes and analysis tools	https://www.htpmd.matr.io , https://github.com/TRI-AMDD/htp_md	272
PPpdb	Database, predictor	Polymer Property Database and Predictor for polymer properties and phase diagrams	https://pppdb.uchicago.edu	
CRIPT	Enabling capability	Community Resource for Innovation in Polymer Technology, graph data model for scalable, efficient and complex polymer data structure	https://criptapp.org	273
SMILES	Enabling capability	Line notation, capable of representing the atomic connectivity of polymers	NA	255
CurlySMILES	Enabling capability	Line notation and tools, handling stereogenicity, electron delocalization charges, extramolecular interactions and so on	https://www.axeleratio.com/csm/py/code/downloads.htm	274
BigSMILES	Enabling capability	Line notation and tools, handling the stochastic nature of polymers	https://olsenlabmit.github.io/BigSMILES	275
G-BigSMILES	Enabling capability	Generative BigSMILES, line notation and tools, generating realistic polymer ensembles	https://github.com/InnocentBug/bigSMILESgen	276
polyDAT	Enabling capability	Generic schema, handling chemical information, synthetical pathways and processing procedures	https://olsenlabmit.github.io/BigSMILES	277
polyGNN	Enabling capability, software	Multigraph neural network, processing polymer repeat units as graphs to map onto properties, developing predictive models	https://github.com/Ramprasad-Group/polygnn	164
polyBERT	Enabling capability, software	Chemical fingerprinting capability, processing polymer chemical structure to learn and predict properties	https://github.com/Ramprasad-Group/polyBERT	33
MaterialsBERT	Enabling capability, software	Large language model, fine-tuned PubMedBERT on 2.4M materials science abstracts, used to extract data available in Polymer Scholar	https://huggingface.co/pranav-s/MaterialsBERT	192
RadonPy	Software, open	Open Python library, automating polymer property calculations using MD simulations	https://github.com/RadonPy/RadonPy	264
Polymer Genome	Predictor	Informatics platform, offering 3 dozen polymer property predictors	https://www.polymergenome.org	24,25
PolymRize	Enabling capability, predictor, commercial	Commercial platform to train models, predict property, and screen for target polymers and formulations	https://polymrize.matmerize.com	

MD, molecular dynamics; NA, not available.

functional theory (DFT)^{84,85}, in conjunction with physical experiments. Early efforts to define screening criteria emphasized the simultaneous attainment of a large electronic bandgap and a high dielectric constant as crucial to achieving high energy density^{55,56}. However, the screening strategy has since been rethought in light of the inverse relationship between these two properties, particularly pronounced in polymers where dielectric response is dominated by electronic polarization.

To elucidate this inverse correlation between bandgap and dielectric constant, high-throughput DFT computations⁵⁵ were conducted for

hundreds of hypothetical polymers constructed from common building blocks found in known polymers^{55,56} (Fig. 2b). The pragmatic choice to focus on polymers with a moderate bandgap (>4 eV) and a moderate dielectric constant (>3), rather than much higher values of both properties, resulted in the discovery of numerous polymers – including a polythiourea, a polyurea and a polyimide (inset, Fig. 2b) – with energy densities exceeding 9 J cm⁻³, twice that of BOPP, or better.

Enhancing the energy density of capacitors at elevated temperatures remains a challenge, whose solution is crucial for not only

substantial space and weight savings but also enabling high-temperature operations. Commercial materials such as BOPP rapidly decline in performance and energy density as temperatures rise (Fig. 2c). To discover high-energy-density polymers capable of withstanding temperatures up to 200 °C, an additional critical criterion is needed to complement those based on bandgap and dielectric constant. This third screening criterion emphasizes that the glass transition temperature must surpass a specified threshold – 200 °C in this case – for stable operation at elevated temperatures. Thus, the new screening criteria to design polymers with high energy density, tolerant to large electric field and temperature, are based on the simultaneous maximization of bandgap, dielectric constant and glass transition temperature.

Following the strategy outlined in Box 1, the most reliable available data (measured or generated using DFT) for these three properties have been used to build robust ML models and create new polymers with substantial energy density, stable across a broad temperature

range up to 200 °C (refs. 7,55,56,80). Using VFS, over 50,000 candidate polymers were virtually generated, starting from suitable commercially available monomers and the ring-opening metathesis polymerization (ROMP) template. Bandgap, dielectric constant and glass transition temperature of the polymers were predicted, and those satisfying the screening criteria specified in Fig. 2c, about 30 polymers, were presented to synthetic chemists. Five of these were chosen, and four were successfully synthesized and characterized experimentally. All four polymers substantially surpassed BOPP, but one of them, a previously unknown polynorbornene dielectric named PONB-2Me5Cl, was a clear outlier with extraordinary energy density over a broad range of temperatures⁸⁰. At 200 °C, PONB-2Me5Cl has an unprecedented energy density of 8.3 J cm⁻³, over an order of magnitude higher than any commercial alternative⁸⁰. The reason for the superior performance of PONB-2Me5Cl is that it simultaneously displays high bandgap, high dielectric constant and high glass transition temperature owing to,

Box 1 | An AI-based, use-inspired and application-driven strategy for polymer design

Application-driven strategies for polymer design share certain key steps.

Define screening criteria

First, a set of screening criteria specified in terms of property values desired for the application must be defined.

Define the search space

Next, a protocol to create a candidate list of polymers must be developed. Although numerous enumerative and generative approaches have been used, the greatest barrier has been to produce materials designs that are genuinely synthetically accessible, cost-effective and safe. A powerful approach proposed recently is ‘virtual forward synthesis’, or VFS, which starts with commercially available monomer molecules and creates polymers using known polymerization reaction templates. Polymers generated using VFS have a naturally high probability of synthetic success.

Machine-learning model development

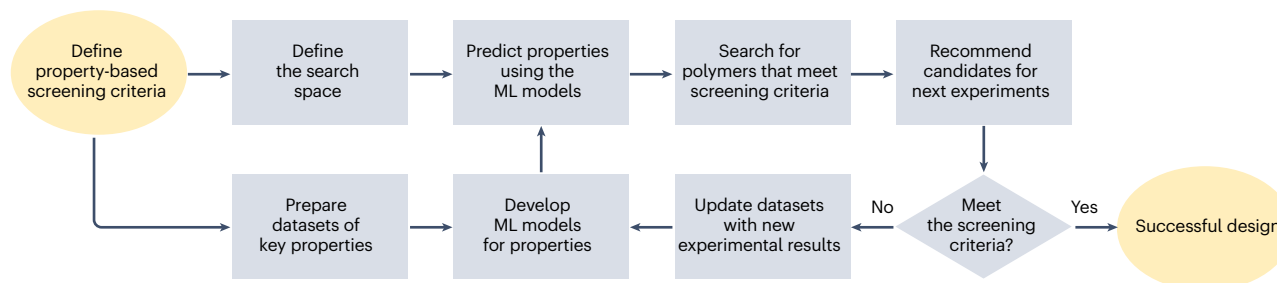
In parallel, reliable machine learning (ML) models are developed to predict the relevant properties rapidly and accurately. ML models

are needed because determination or estimation of most properties of new-to-the-world polymers using traditional options is too slow (for example, physical experiments), impractical (for example, simulations based on density functional theory) or semiquantitative at best (for example, classical simulations via molecular dynamics simulations). Developing these ML models requires a sufficiently large and diverse initial training dataset, produced using prior physical experiments or computational methods. The polymer property datasets are then converted to machine-readable numerical form (or ‘fingerprinted’), followed by ‘learning’ the mapping between polymers and properties using suitable ML algorithms.

Candidate selection and recommendation

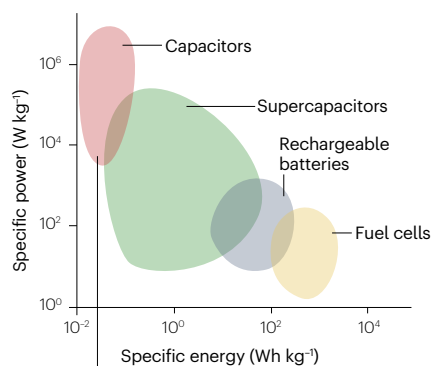
Finally, properties of relevance may be predicted using the ML models developed previously for the generated list of polymers, and those that meet the screening criteria are selected and recommended for physical experimentation and validation. The fresh data thus obtained from physical experiments, whether meeting the required criteria or not, may be used to restart the design cycle, which may progress in an iterative manner (also referred to as ‘active learning’) until the design goals are reached.

An AI-based application-specific polymer design strategy



AI, artificial intelligence.

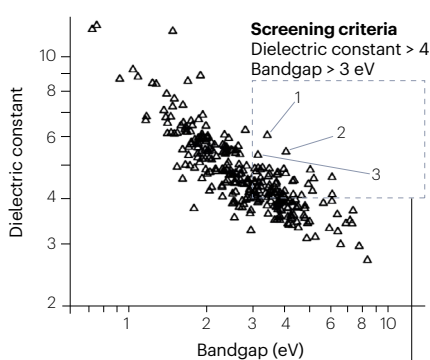
a Ragone plot of energy storage systems



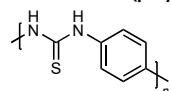
Applications of polymer dielectric capacitors

- Hybrid and electric vehicles
- Pulsed power systems
- Wind turbine generators
- Aircraft launchers
- Aerospace applications
- Fusion power and technologies

b Computational screening for polymer dielectrics



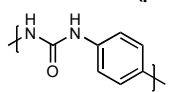
1: NH-CS-NH-C6 (polythiourea class)



Dielectric constant
DFT: 6.7
Experimental: 6.0

Bandgap
DFT: 3.1 eV
Experimental: 2.5 eV

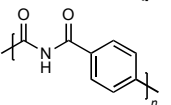
2: NH-CO-NH-C6 (polyurea class)



Dielectric constant
DFT: 5.8
Experimental: 4.9

Bandgap
DFT: 3.9 eV
Experimental: 3.4 eV

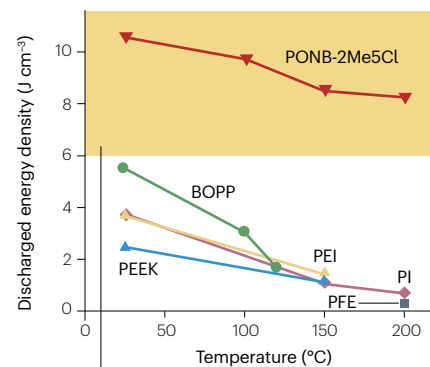
3: CO-NH-CO-C6 (polyimide class)



Dielectric constant
DFT: 4.8
Experimental: 4.9

Bandgap
DFT: 4.0 eV
Experimental: 3.3 eV

c Designed high-temperature polymer dielectrics



Property	Desired value
Dielectric constant	>3
Bandgap	>3 eV
Glass transition temperature	>150 °C

PNB and PONB polymers

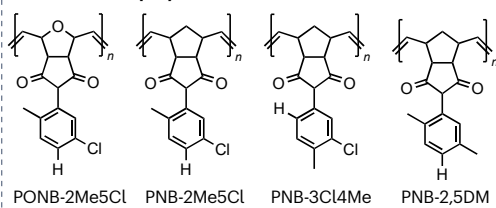


Fig. 2 | Dielectric polymers for energy storage. **a**, Ragone plot of various classes of energy-storage systems, including electrostatic capacitors. **b**, The inverse relationship between dielectric constant and bandgap, which defines the screening for optimal candidates. Three candidates, labelled 1, 2 and 3, were selected, synthesized and validated. **c**, PONB-2Me5Cl, a newly discovered and tested polymer for energy storage at high temperature, aided by a rethought set of screening criteria (inset table) to address high-temperature behaviours. Four candidate polymers in this class were synthesized and tested; all of them

were high performing, with PONB-2Me5Cl displaying the best performance, substantially higher than regular polymers like biaxially oriented polypropylene (BOPP), polyether ether ketone (PEEK), polyetherimide (PEI), polyfluoroethylene (PFE) and polyimide (PI), by up to one order of magnitude, especially at high temperatures. DFT, density functional theory; PNB, polynorbornene; PONB, polyoxanorbornene. Panel **b** reprinted from ref. 56, Springer Nature Limited. Panel **c** reprinted from ref. 80, Springer Nature Limited.

respectively, the lack of conjugation along the backbone, the rotatable polar group in the pendant side chain, and the stiff backbone combined with a bulky side chain. Such a combination of features and properties is rare (and possibly non-existent) in polymers synthesized thus far, highlighting how AI algorithms can aid in extending discoveries beyond conventional human imagination.

Looking ahead, there are additional opportunities to explore. Whereas the work discussed above revolved around a specific organic polymerization template (namely ROMP), there are hundreds of other templates available^{51–53}, some even incorporating metal atoms in the backbone that can substantially increase dielectric constant^{86,87}. Each template may be coupled with available (in orders of billions) and new-to-the-world (countably infinite) monomers, which could lead to numerous hypothetical, but synthesizable, polymers that are even better than the known and discovered candidates. However, practical considerations, such as monomer cost, complexities and scalability of polymerization processes, toxicity concerns, the role of

polymer–electrode interfaces^{88,89}, and aspects related to recyclability and sustainability, impose limitations.

Polymers for fuel-cell applications

Fuel cells are devices that generate electricity directly from the chemical energy of reactants, namely a fuel (such as hydrogen) and an oxidant (such as oxygen)⁹⁰. A typical fuel cell consists of a fuel electrode (anode), an oxidant electrode (cathode) and an electrolyte filled in between (Fig. 3a). The electrolyte is a material that allows the charge carriers (such as protons) to transport efficiently while blocking carriers of opposite charge (such as electrons) and the gas reactants from penetrating and diffusing. The catalyst layers, where the oxidation and reduction reactions occur, are typically created by binding nanoparticles of electrocatalysts (such as platinum) to a support with a polymeric ionomer solution. The electrolyte and the catalyst layers of both electrodes accommodate all the essential chemical reactions and charge transports. Starting from early concepts in the 1840s, fuel cells have

now been used in transportation, consumer electronics, residential power supply and more^{91–94}. Compared with capacitors, fuel cells are generally higher in energy density but lower in power density (Fig. 2a).

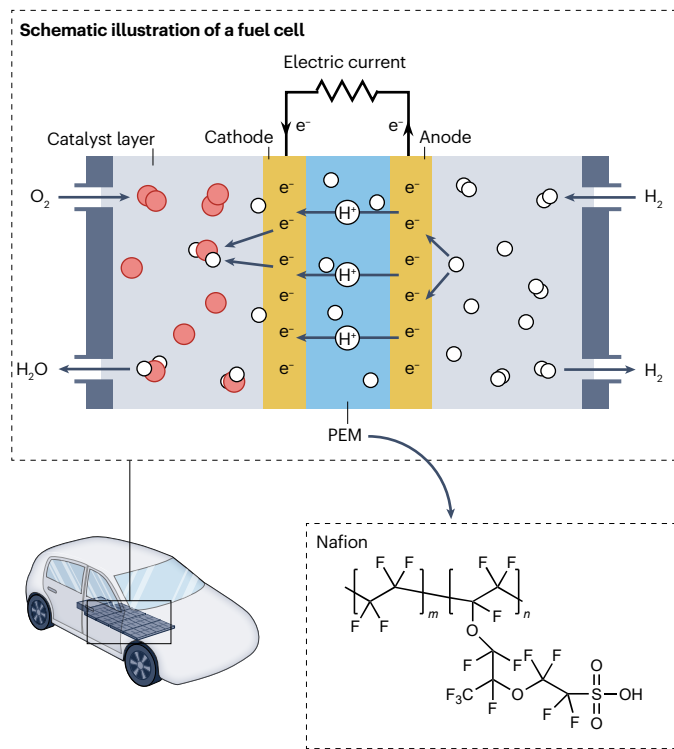
Nafion, a perfluorosulfonic acid (PSFA) polymer (chemical structure in Fig. 3a), is the currently dominant proton exchange membrane (PEM) – that is, the proton-conducting electrolyte – and ionomer of both electrodes in modern fuel cells^{8,91–94}. Given the conflicting requirements for PEM and ionomers⁵⁷ (Fig. 3b), this choice is not always optimal. For example, the low permeability of Nafion to O₂ (refs. 95,96) is good for a PEM but not for a cathode ionomer^{97–99}. In addition, the required proton conductivity of Nafion can only be obtained when it is submerged in water or when its humidity is nearly 100%, a challenging working condition to maintain. Moreover, the relatively low glass transition temperature ($T_g \approx 120$ °C) of Nafion limits its working temperatures. Finally, Nafion is expensive. These factors, among others, drive the search for Nafion alternatives^{100–106}, specifically fluorine-free materials¹⁰⁵. The main approaches used thus far are empirical, focusing on controlling certain key features of Nafion, such as the sulfonic (–SO₃) group¹⁰⁵, and exploring its related chemistries^{97,100}, with limited successes.

In fuel-cell design, ML approaches have been used mostly for device modelling^{100,107–109}. In a rare work⁵⁷ using an ML strategy to discover new polymers for fuel-cell applications, a list of screening criteria were established in terms of important properties of PEM and ionomers (Fig. 3b). Suitable datasets were curated, and ML models needed for the

properties were developed (some are available in [Polymer Genome](#)). The most important model was trained concurrently on two datasets of proton conductivity and water uptake, enabling it to predict these correlated properties simultaneously. This model is an example of the multitask learning technique, used to fuse multiple data channels as elaborated in the ‘Computational–experimental data fusion and multifidelity learning’ section. In this design problem, the VFS approach was applied in a restricted manner by considering about 60,000 homopolymers and copolymers that were experimentally synthesized and reported. More than 60 polymers were identified as possible candidates for PEM, cathode ionomer and anode ionomer (examples are shown in Fig. 3c).

Future work could address some critical gaps in this initial attempt towards designing polymers for fuel-cell applications. First, owing to the historical emphasis on PSFA membranes, the curated data are dominated by polymers with sulfonic ($-\text{SO}_3$) groups, crucial for water retention in PSFA (although other groups may lead to similar functionality)⁵⁷. Thus, a criterion of ‘having the sulfonic group’ may be used to narrow down the candidate pool. Second, this work was limited in the search space definition, containing only previously reported polymers (albeit for any applications, not just for fuel cells), further restricting the number of discoveries. Going forward, when the training data contain other polymer classes and when the full power of VFS is exploited to cover the vast space of synthesizable polymers^{51,52}, numerous highly qualified candidates for fuel-cell PEM and ionomers can be expected. Development may also be needed to address

a Fuel cell concept and applications



b Key properties required for fuel-cell applications

Key properties	Desired values		
	PEM	Cathode ionomer	Anode ionomer
Proton conductivity (S cm ⁻¹)	High	High	High
O ₂ permeability (barrer)	<18	>18	NA
H ₂ permeability (barrer)	<37	NA	>37
Bandgap (eV)	>4	NA	NA
Glass transition temperature (°C)	>123	>123	>123
Thermal decomposition temperature (°C)	>373	>373	>373
Young's modulus (MPa)	>156	>156	>156
Having -SO ₃ group?	Yes	Yes	Yes

C Some candidate polymers for PEM

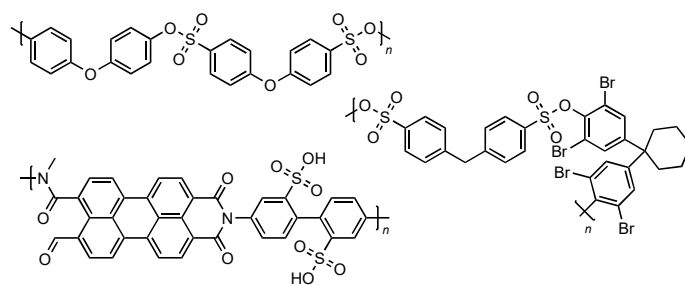


Fig. 3 | Polymers for fuel cells. **a**, A schematic illustration of a fuel cell, which uses Nafion for a proton exchange membrane (PEM) and ionomers. **b**, Screening criteria suggested for a PEM, cathode ionomer and anode ionomer. **c**, Some polymers discovered for PEM.

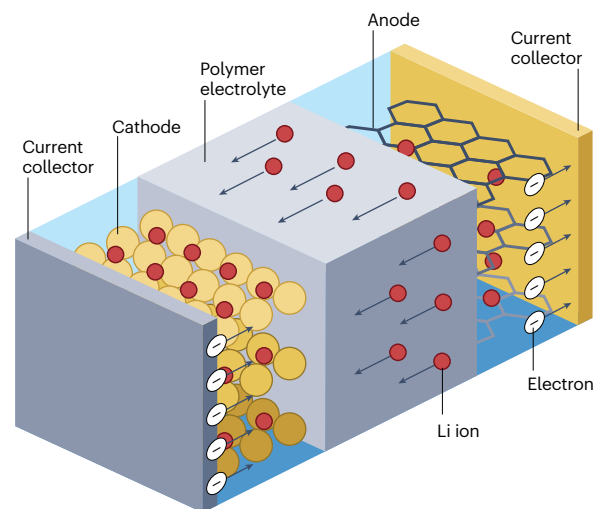
concerns such as the toxicity and sustainability of the candidates, and the scalability of their synthetic routes. As next-generation fuel cells that involve anions rather than protons are being considered^{110,111}, the current approach will need to expand further to meet the requirements of anion exchange membrane designs.

Polymers for Li-ion battery electrolytes

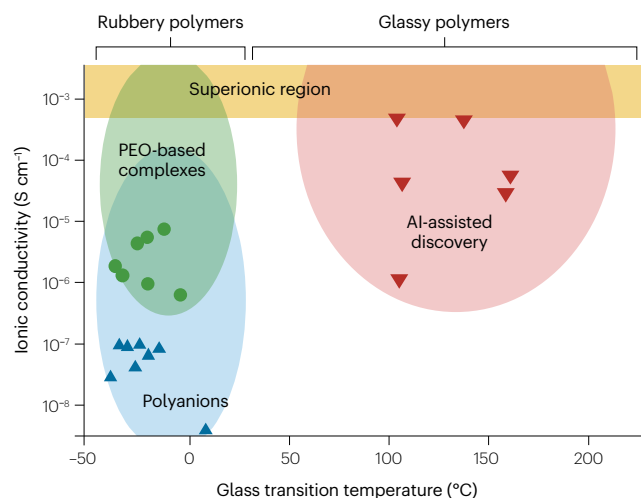
Lithium-ion batteries are currently used in almost all kinds of electrically powered devices, from portable electronics to hybrid cars, electric cars and aeroplanes, and their adoption is set to surge as global electrification progresses. Serious efforts to develop Li-ion batteries

commenced in the 1960s, then accelerated owing to the oil crisis of the early 1970s¹¹². The main advantages of Li-ion batteries are high densities of energy and power, robustness and long life cycle. In the Ragone plot (Fig. 2a), Li-ion batteries are intermediate to capacitors and fuel cells. A Li-ion battery cell (Fig. 4a) has an anode (the reductant) and a cathode (the oxidant), where Li ions are deposited and released, and an electrolyte, through which Li ions are transported. A separator is needed to prevent physical contact between the electrodes while allowing Li ions to shuttle through¹¹³. Optimizing the current materials and discovering new ones for anodes^{114,115}, cathodes^{116–119} and, especially, electrolytes^{58–61,119}, have been important foci of the field^{120–122}.

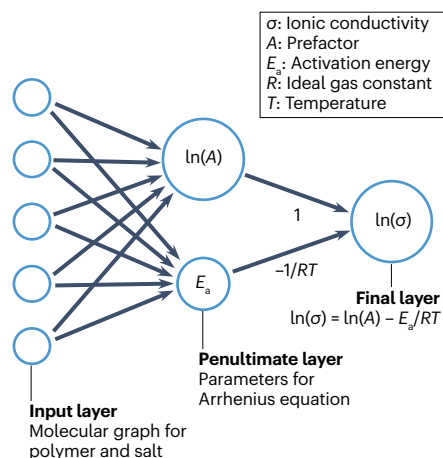
a Lithium-ion battery concept (discharging status)



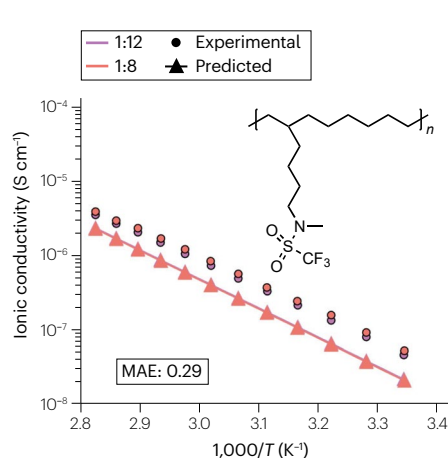
b AI-assisted discoveries of polymers for electrolytes



c A chemistry-informed neural network



d Arrhenius plot of ionic conductivity



e Discovered polymer electrolytes

Polymer 2D structure	Ionic conductivity (mS cm⁻¹)
	2.041 ± 0.384
	1.605 ± 0.054
	1.517 ± 0.101
	1.515 ± 0.199
	1.501 ± 0.124
	1.491 ± 0.176
	1.481 ± 0.262
	1.473 ± 0.140

Fig. 4 | Polymers for Li-ion batteries. **a**, A schematic structure of a Li-ion battery. **b**, Some artificial intelligence (AI)-assisted discoveries of solid polymer electrolytes (SPEs) compared to previously studied SPEs, and their ionic conductivity relationship with the glass transition temperature. **c**, ChemArr, a physics-informed neural network, enforcing the Arrhenius formula in the predictions of Li-ion conductivity. **d**, Predicted Li-ion conductivity as a function

of temperature, which agrees well with the Arrhenius formula. **e**, Some polymers discovered for Li-ion battery electrolytes using machine learning strategies. MAE, mean absolute error; PEO, poly(ethylene oxide). Panel **b** reprinted with permission from ref. 146, American Chemical Society. Panels **c** and **d** reprinted from ref. 143, CC BY 4.0.

The search for new electrolyte materials is motivated by safety concerns and the flammability of current liquid electrolytes, such as solutions of lithium hexafluorophosphate and some flammable organic liquids. Such materials are believed to lead to spontaneous explosions and fires in some Li-ion battery units¹²³, as in more than 20 documented fires in Tesla models since their introduction¹²⁴ and a Boeing 787 Dreamliner aeroplane¹²⁵. As a potential alternative, solid polymer electrolytes (SPEs) that are light, cheap and, most importantly, safe have been examined, including those based on poly(ethylene oxide) (PEO)^{126–129}, polyacetals^{130,131}, polyethylene carbonates¹³², polyesters¹³³, polyacrylonitrile (PAN)¹³⁴, poly(vinyl alcohol) (PVA)¹³⁵, poly(methyl methacrylate) (PMMA)¹³⁶, and polymer blends such as PEO/PAN¹³⁷ and PVA/PMMA¹³⁸. However, these SPEs still have major disadvantages for use in Li-ion batteries, one of which is their low Li-ion conductivity ($\leq 10^{-5}$ S cm⁻¹) at practical operating temperatures (the target conductivities are $\geq 10^{-4}$ S cm⁻¹).

Although a number of SPE candidates have been explored, they represent only a tiny fraction of the polymer space, which comprises tens of thousands of already known and synthesized polymers^{25,57}, or millions of commercially available or easy-to-synthesize monomers that can be polymerized via hundreds of known polymerization reaction templates. There is a good explanation for this: putting physical experimentation aside, physics-based computational methods are not ready to evaluate SPE candidates. Classical molecular dynamics (MD) simulation, the most practical computational method to estimate Li-ion conductivity today in polymers^{128,129,139,140}, is, at best, semiquantitative, and requires extreme care and specialized skills. Good candidates for potential SPEs are likely to be somewhere in the vast untapped polymer space, awaiting discovery and deployment.

ML approaches have made some progress^{42,139,141–144}, especially in accelerating the MD-based evaluations of SPEs and decision-making procedures. Bayesian optimization has been used to drive coarse-grained MD explorations of the polymer space. One endeavour¹⁴² identified polymer blends with optimized Li-ion conductivity and mechanical strengths; another⁴², by generating a big volume of data, uncovered the relationships between the Li-ion conductivity and relevant atomic-level features such as molecule size and non-bonding interaction strengths. Based on an interesting idea to ‘accelerate’ the MD simulations, a ML capability was developed¹³⁹ to perform early predictions of the equilibrium Li-ion transport properties of a polymer from its chemo-structural descriptors and information obtained within the first 0.5 ns of the MD trajectory. This scheme could reduce the MD simulation times by 90%, substantially accelerating explorations for SPEs¹³⁹. In recognition of the importance of high-fidelity data in the field, a cloud-based platform was established¹⁴⁵ to share raw data from 6,286 MD trajectories of amorphous polymer electrolytes and standard post-processing and analysis tools.

The impact of ML approaches extends beyond the acceleration of MD simulations. By constructing a database of SPEs, a transfer-learned graph neural network was trained and used¹⁴⁶ to search over 9,600 combinations of polymers, dopants, salts and other parameters, leading to the discovery of eight polyphenylene sulfides, which were then validated experimentally (six of them are shown in Fig. 4b). ChemArr is a physics-informed neural network in which the Arrhenius equation, which governs the temperature dependence of the Li-ion conductivity, is explicitly encoded¹⁴³ (Fig. 4c). The model was trained on a dataset of 7,133 experimental Li-ion conductivity data points curated for 247 unique polymers, and its power to predict Li-ion conductivity was demonstrated on two unseen new polymers, named P_CODC₄CF₃SA

(Fig. 4d) and P_C₁₀PA_MC. Attempts to design new SPEs were further extended to involve a quantum annealer, inverting the developed regression model to identify the ‘ideal features’ of the desired SPEs¹⁴⁷. Existing databases were then searched, uncovering a trithiocarbonate-based polymer resembling the ideal SPE. This polymer was synthesized and shown to offer a conductivity of 10^{-6} S cm⁻¹ and thermal stabilities above 80 °C. Efforts in the past 2–3 years have leveraged the development of generative models (Generative Pre-trained Transformer (GPT)-based and diffusion-based) to conditionally and continually design new homopolymers with high predicted Li-ion conductivities^{148,149}. Using this approach, 19 polymer repeat units were found¹⁴⁹ to display computed ionic conductivities (via MD simulations) surpassing that of PEO (Fig. 4e).

Although Li-ion conductivity is the most important property of a SPE for Li-ion batteries, SPEs should also have a large electrochemical stability window, which controls the open-circuit voltage and ultimately the cycle life of the batteries, and should be mechanically strong, thermally stable and durable for safety reasons. The critical gaps to address are the development of necessarily bigger, more diverse, high-quality databases, and training powerful predictive models of the desired properties. The quantitative screening criteria for SPEs, like those in Fig. 2c and Fig. 3b, as well as for the novelty and validity of the candidates¹⁴⁸ should be established. Then, the vast space of synthetically accessible polymers can be screened to identify superior SPE candidates. Regardless of the specific approaches adopted, care must be taken to safeguard the likelihood of finding viable and scalable synthetic routes of the SPE candidates.

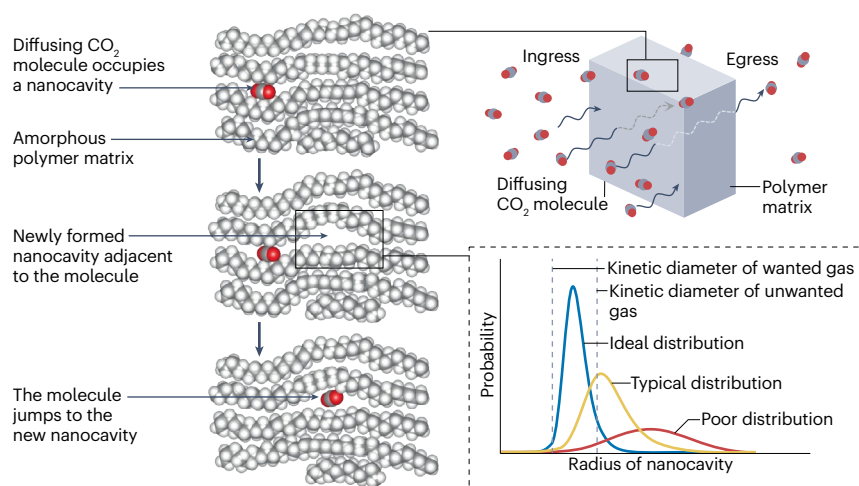
Membranes for gas separation

Using synthetic polymer-based membranes to separate gas mixtures^{10,62,150–157} – for example, removing CO₂ from natural gas or removing O₂ from air – is favoured over competing technologies owing to the suitable combinations of energy efficiency, cost and size in these membranes¹⁵⁸. Compared with distillation, which conventionally requires a massive amount of heat, membranes can, in principle, separate gas mixtures in the presence of just a pressure gradient (Fig. 5a). An important performance measure of a gas separation membrane is the selectivity, which, for binary mixtures, is defined as $\alpha_{ij} \equiv P_i/P_j$, where P_i and P_j are the permeabilities of i , the more permeable gas, and j , the less permeable gas. Although many other technological factors influence the success or failure of a particular membrane, an ideal membrane material will have high selectivity and high permeability to the wanted gas. This combination simplifies the membrane engineering process and reduces the operating and capital costs.

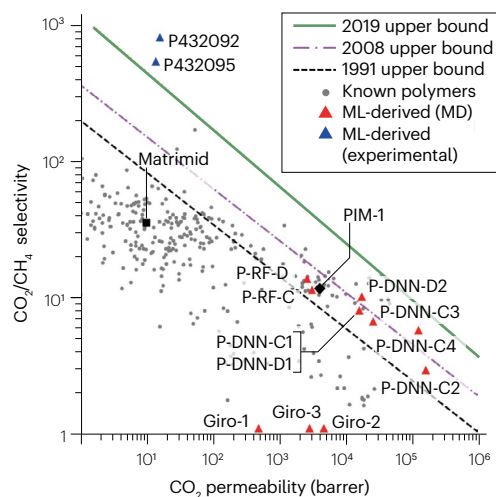
The selectivity and permeability depend heavily on the size of ‘free volume elements’ (FVEs) – the small, often-ephemeral gaps between polymer chains – and the frequency at which these gaps appear and disappear owing to thermal fluctuations in the polymer chains. The selectivity is maximal when the FVEs have a uniform size, preferably positioned between the kinetic diameters of the desired and undesired gases^{159,160}. Therefore, an ideal FVE size distribution should be tight and appropriately centred (inset, Fig. 5a). The permeability is high when FVEs with appropriate size for gas diffusion are created at a rapid frequency. Typically, materials that exhibit high frequencies of FVE creation often have a broad distribution of FVE sizes, giving rise to the well-known trade-off between permselectivity and permeability^{157,159} (see Fig. 5b for an example).

The most obvious consequence of the trade-off is the presence of a performance upper bound, pointed out in 1991 (refs. 157,159).

a Gas atom transport in polymer membranes

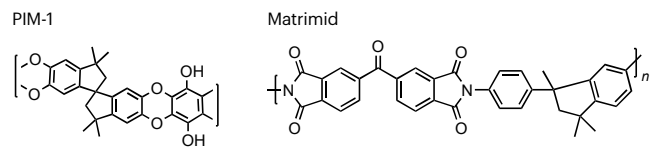


b Robeson plot for gas separation membranes



c Some notable and newly discovered membranes for gas separation

Known polymers



ML-derived polymers

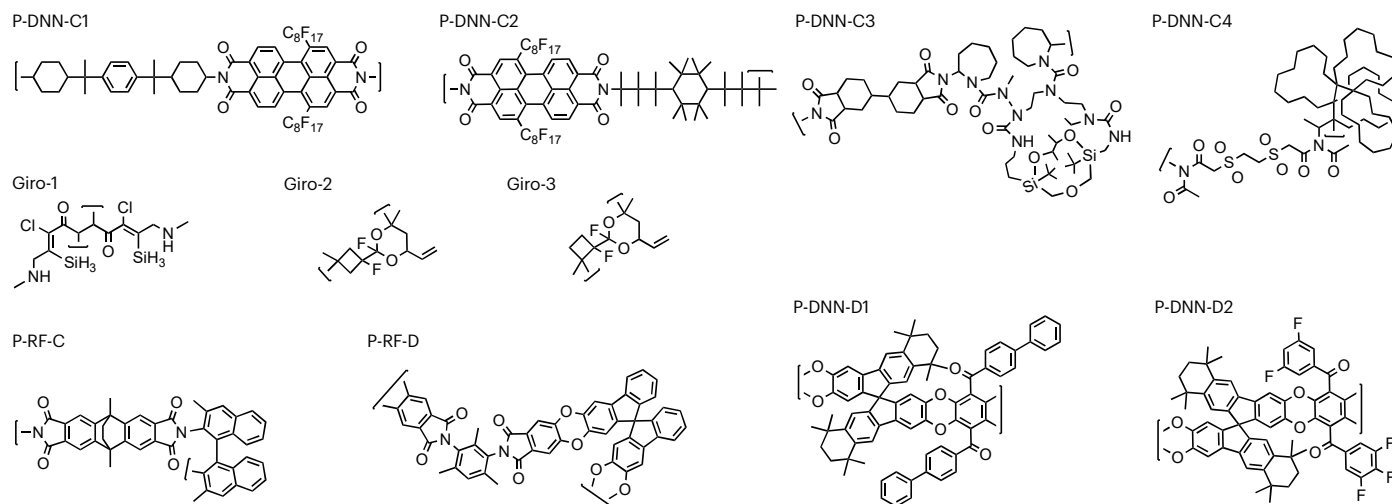
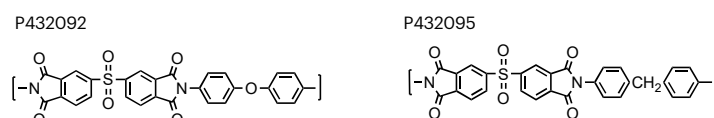


Fig. 5 | Polymers for the separation of gas mixtures. a, Diagram of a gas separation membrane, working in the presence of a pressure difference between ingress and egress, and three pore-size distributions that are ideal, normal and poor for the selectivity. **b**, Robeson plot for CO₂/CH₄ selectivity, given with respect to the CO₂ permeability, showing the trade-off between permeability and

selectivity, how this trade-off upper bound has advanced over the years, and some classic membrane polymers discovered over the years. **c**, Chemical structures of some known and machine-learning (ML)-derived polymers. For ML-derived polymers, data were generated either by molecular dynamics (MD) simulation or experiments.

Residing below this bound are the known polymers of that time, including Matrimid, a polyimide membrane that is still commercially used today^{161,162}. Since then, a handful of new polymers have been discovered that surpass the 1991 bound, establishing two new bounds dated in 2008

and 2019 (refs. 147,160). One of the earliest discoveries that pushes the 1991 bound was PIM-1, or Polymer of Intrinsic Microporosity 1, which possesses a new repeat unit chemistry¹⁴⁷. Specifically, PIM-1 and other PIMs feature a site of contortion (that is, a spiro centre) in every repeat

unit that ‘kinks’ the polymer chain at extreme angles. This feature alone is often insufficient to make a polymer a ‘PIM’; another salient design motif is the incorporation of the spiro centre into a ladder polymer. The combination of these features, unseen in previous polymers, results in the record-breaking permeability of PIM-1. In summary, new chemistries can substantially improve performance, continuously pushing the existing bounds upwards, although this process might have a conceptual upper limit¹⁵³.

Although permeability and selectivity may be estimated using MD simulations, this method is not quantitative enough to discover new and improved gas separation membranes. Works in the past 6–7 years have shifted towards ML approaches, training predictive models on past data to estimate the gas permeability from the chemical structure^{25,154,155,163–165}. In 2022, six ML models for the permeabilities of He, H₂, O₂, N₂, CO₂ and CH₄ were developed and used¹⁵⁵ to screen more than 9 million hypothetical polymers, identifying thousands predicted to be ultrapermeable to O₂ and CO₂. Eight candidates, named P-DNN-C1 through P-DNN-C4, P-DNN-D1, P-DNN-D2, P-RF-C and P-RF-D (Fig. 5c), were validated using MD simulations. In another effort, six models for the permeability of the six gases mentioned above were developed and used¹⁵⁴ for 11,325 polymers from the National Institute for Materials Science (NIMS) database, predicting hundreds lying above the upper bounds for the O₂/N₂ and CO₂/CH₄ separations. Two of them, namely poly[(1,3-dioxoisindoline-2,5-diyl)sulfonyl(1,3-dioxoisindoline-5,2-diyl)-1,4-phenyleneoxy-1,4-phenylene] and poly[(1,3-dioxoisindoline-2,5-diyl)sulfonyl(1,3-dioxoisindoline-5,2-diyl)-1,4-phenylenemethylene-1,4-phenylene] (labelled as P432092 and P432095 in the NIMS database), were synthesized and tested for CO₂/CH₄ separation (Fig. 5c). Targeting CO₂ separation from N₂ for carbon capture, three models, including one for \mathbb{P}_{CO_2} were developed and used to identify hundreds of high- \mathbb{P}_{CO_2} polymers¹⁶³. Three of them, labelled as Giro-1, Giro-2 and Giro-3 (Fig. 5c), display high \mathbb{P}_{CO_2} , as estimated by MD simulations.

These exemplary works highlight the role of ML approaches in designing membranes for gas separation. Yet critical challenges remain. One persistent challenge, common to all application domains, is ensuring that the recommended polymers are synthetically feasible and scalable. As an illustration, only two of the ML-derived polymers shown in Fig. 5c have been synthesized; the remaining candidates are difficult to make. Constructing a vast space of synthetically accessible polymers using VFS could be a solution, provided VFS includes reaction templates and chemistries with ladder features and spiro centres.

A second challenge, also relevant to all applications, is that the models need to train on data spanning large enough chemical spaces. To alleviate this challenge, MD-simulated data (although low in fidelity) for \mathbb{P}_i (ref. 166) and related or correlated properties may be generated to augment available measured data. For example, the solubility S_i and diffusivity D_i of a gas in a polymer i are related to the permeability by $\mathbb{P}_i = D_i \times S_i$. Furthermore, the fractional free volume (FFV) of polymers, as discussed above, is strongly correlated to their permeability¹⁶⁷. Data (measured and/or simulated) from any of two or more of these four properties (\mathbb{P}_i , D_i , S_i , FFV), may be leveraged in a multitask ML architecture to learn all properties simultaneously, improving the accuracy and the robustness of the \mathbb{P}_i predictors.

Finally, ‘ageing’ (or degradation) problems that pervade gas separations applications should be addressed. Over time, \mathbb{P}_i can decrease owing to altered or degraded distribution of FVEs, and this behaviour should be managed appropriately. Membranes for gas separation should be mechanically, thermally and chemically stable over 5+ years for viable real-world applications. Testing these long time frames in an experimental

laboratory is typically infeasible; development of algorithms to predict ageing of a given polymer would accelerate development in this field. Beyond membrane performance and ageing, incorporating additional properties in the screening criteria relevant for scale-up such as tensile modulus, glass transition temperature and thermal decomposition temperature would be impactful.

Membranes for organic liquid mixture separations

Polymer-membrane-based separations of non-aqueous or organic–water liquid mixtures solve a different class of problems relative to gas separations. Membrane-based separations of liquid mixtures, driven by pressure rather than heat, are energy and economically efficient^{63–66}. This method is important to the chemical and pharmaceutical industries, where separation processes, such as in the recovery of organic solvents, could account for up to 40–70% of the capital and operating cost⁶³. Technically, membranes can be used to separate organic compounds with similar boiling points⁶⁵ or that are temperature sensitive¹⁶⁸; these types of separations are challenging or expensive to carry out with incumbent technologies, such as vacuum distillation. Important liquid mixture separations accessible to membranes include water purification, solvent recovery, solute concentration, diluent separation, iterative synthesis of oligomers, homogeneous catalyst recovery, natural product extraction, membrane reactors and solute fractionation, among many others^{63–65,169}. Membranes that can separate the liquid phase of small molecules, such as ethanol and iso-octane, typically operate in a solution-diffusion regime, like gas separations.

One difficulty for this class of membrane-based separations is estimating how well the membrane will perform when challenged with a new complex mixture. The efficiency of a separation depends on several factors, including the characteristics of the liquid mixture to be separated (such as the number of distinct solute and solvent types, concentration, size, polarity and so forth of each solute and solvent), the choice of the membranes, the operating conditions (such as pressure) and the time-dependent performance fluctuations (that is, the ageing). In the design problem of membranes for separation of organic liquid mixtures specifically, this high dimensionality of the search space makes it a daunting task for traditional physics-based models alone.

To handle complex mixtures, ML models based on neural networks, random forest models and support vector machines were established to predict important liquid mixture transport properties such as the permeance and rejection of mixtures containing a solvent and a solute¹⁶⁹. Nonlinear regression techniques, whose parameters are determined by a combination of genetic programming and global deterministic optimization, were used to predict the permeance of pure solvents and solvent mixtures through membranes¹⁷⁰, and the solute rejection in liquid mixtures containing multiple solvents and/or solutes¹⁷¹. Building on this work, some ML models were developed¹⁷² that can predict the solute rejection using the molecular structure of the solute as input. Unlike previous counterparts that can handle a few fixed solutes, these models can be generalized to any solute.

Each of the above works are suitable for mixtures containing solutes of a particular size (100–2,000 g mol^{−1}) and often much smaller solvents that permeate via a pore-flow style mechanism. However, these models are unsuitable for mixtures of small molecules, or complex mixtures (for example, crude oil containing thousands of components), with near equal concentrations, in which there is no clear solvent or solute. Moreover, not all membranes operate with a pore-flow transport modality; indeed, many effective membranes

```
graph LR; Input[Input] --> GenParam[Generation of parameters]; GenParam --> TransMod[Transport modelling]; TransMod --> Output[Output]
```

The flowchart illustrates the proposed solution-diffusion model for gas transport in polymer membranes, consisting of four sequential stages:

- Input:**
 - Polymer structure
 - Solvent structure
 - Solvent activity
- Generation of parameters:**
 - Solvent-polymer diffusivities using ML model
 - Solvent-polymer sorption parameters using ML model
 - Solvent molar volumes
 - Solvent vapour pressure
 - Solvent Hansen solubility parameters
 - Pressure
 - Temperature
 - Composition of feed
 - Thickness of membrane
- Transport modelling:**
 - Solving Maxwell-Stefan diffusion equation for solution-diffusion membranes
- Output:**
 - Partial fluxes
 - Composition of permeate

Matrimid

Chemical structure of Matrimid polymer repeat unit. It consists of a 4,4'-oxydiphenyl group connected to two phthalimide rings. The polymer backbone is formed by the phthalimide rings, with the 4,4'-oxydiphenyl group acting as a side chain. The repeat unit is shown with a subscript n .

DUCKY-9

Chemical structure of DUCKY-9 polymer repeat unit. It features a fluorene core substituted with a methoxy group and a 4,4'-oxydiphenyl group. The polymer backbone is formed by the fluorene units, with the 4,4'-oxydiphenyl group acting as a side chain. The repeat unit is shown with a subscript n .

DUCKY-10

Chemical structure of DUCKY-10 polymer repeat unit. It features a fluorene core substituted with a methoxy group and a 4,4'-oxydiphenyl group. The polymer backbone is formed by the fluorene units, with the 4,4'-oxydiphenyl group acting as a side chain. The repeat unit is shown with a subscript n .

Torlon

Chemical structure of Torlon polymer repeat unit. It consists of a 4,4'-oxydiphenyl group connected to two phthalimide rings. The polymer backbone is formed by the phthalimide rings, with the 4,4'-oxydiphenyl group acting as a side chain. The repeat unit is shown with a subscript n .

SBAD-1

Chemical structure of SBAD-1 polymer repeat unit. It features a fluorene core substituted with a methoxy group and a 4,4'-oxydiphenyl group. The polymer backbone is formed by the fluorene units, with the 4,4'-oxydiphenyl group acting as a side chain. The repeat unit is shown with a subscript n .

Tested on Torton
12-component
60 bar
295 K

Total flux ($\text{l m}^2 \text{h}^{-1}$)
Measured: 0.004
Predicted: 0.019 ± 0.013

Tested on Matrimid
9-component
40 bar
295 K

Total flux ($\text{l m}^2 \text{h}^{-1}$)
Measured: 0.81
Predicted: 0.69 ± 0.56

Tested on DUCKY-9
9-component
40 bar
295 K

Total flux ($\text{l m}^2 \text{h}^{-1}$)
Measured: 4.65 ± 0.14
Predicted: 2.10 ± 1.16

Tested on DUCKY-10
9-component
40 bar
295 K

Total flux ($\text{l m}^2 \text{h}^{-1}$)
Measured: 8.22 ± 0.23
Predicted: 3.33 ± 1.61

Tested on SBAD-1
Boiling point range (°C)

- <100
- 100–200
- 200–300
- 300–400
- 400–500
- 500–600
- >600

Total flux ($\text{l m}^2 \text{h}^{-1}$)
Measured: 0.85 ± 0.35
Predicted: 0.88 ± 0.45

Tested on DUCKY-9
Boiling point range (°C)

- <100
- 100–200
- 200–300
- 300–400
- 400–500
- 500–600
- >600

Total flux ($\text{l m}^2 \text{h}^{-1}$)
Measured: 1.375
Predicted: 0.34 ± 0.23

Fig. 6 | Polymers for the separation of complex liquid mixtures. **a**, Multiscale data-driven transport modelling framework informed by physics and/or chemistry. **b**, Five polymer membranes used to validate the framework. **c**, Test results from four selected membranes on 9-component or 12-component hydrocarbon mixtures, showing close correspondence between predicted and

measured results. **d**, Test results from two selected membranes on real crude oils with thousands of liquid components, showing close correspondence between predicted and measured results. ML, machine learning. Panels **c** and **d** reprinted from ref. 173, CC BY 4.0.

operate in the solution-diffusion regime. To extend the range of possible mixtures and operation regimes, a multiscale approach integrating a multicomponent mass transport scheme with a physics-informed ML approach was developed¹⁷³ (Fig. 6a). First, a neural network was used to predict the diffusivity and sorption uptake of each individual mixture component for a given membrane. Then, these predictions were fed into a Maxwell–Stefan solution-diffusion transport model¹⁷⁴, from which each component's flux was predicted. Importantly, this approach generalizes to the separation of arbitrary mixtures and linear membranes. Physical experiments then validated the framework's ability to predict the permeation outcomes of complex liquid mixtures through membranes. Five polymers, two of them (Torlon and Matrimid) commercially available and the others (DUCKY-9, DUCKY-10 and SBAD-1) recently lab-synthesized^{66,175} (Fig. 6b), and mixtures of either 9 or 12 liquid components (Fig. 6c) or real crude oils with thousands of liquid components (Fig. 6d), were tested. This multitiered approach could predict the separation of complex mixtures to within 6–7% of the measured values¹⁷³ (Fig. 6c,d).

Although the ability to build robust ML models that can handle multicomponent industrial-scale complex liquid mixtures has been demonstrated, their potential can be expanded by continually training the models on emerging new data for other complex mixtures. Furthermore, the true opportunity is to leverage these models to suggest new polymer membranes and optimal operating conditions for industrially important liquid mixtures. This gap may be addressed with the adoption of generative methods, like the VFS approach, to generate a diverse pool of candidates, forecast permeation performance for complex liquid mixtures and subsequently identify those meeting predefined criteria.

Conducting conjugated polymers

The development of conducting polymers^{176–180} marks a milestone in the history of polymers. In the early 1970s, polyacetylene was synthesized, exhibiting semiconducting behaviour ascribed to the delocalized π -electrons arising from its conjugated structure¹⁷⁹. Treating polyacetylene with Lewis acids or bases was revealed to substantially enhance its conductivity, sometimes by up to 13 orders of magnitude¹⁸⁰, and turning it into a conductor. Because this process involves the removal or addition of electrons to the polymer chains, it is termed 'doping' in analogy to the doping procedures adopted in silicon technology. Doping is important to controlling the conductivity of conducting polymers, allowing them to find applications in organic light-emitting diodes^{181,182}, organic field-effect transistors^{183,184}, organic solar cells^{185,186}, biomedicine^{187–190} and beyond.

Using a curated set of 389 experimental data points covering 226 polymers and 65 dopants, with conductivities spanning 16 orders of magnitude, a data-driven approach was able to accelerate the identification of suitable candidates for conducting polymers¹⁹¹. Classification and regression models to predict conductivity were developed¹⁹¹ from handcrafted chemical fingerprinting schemes. The classification model categorized the conductivity as low, medium or high, while the regression model provided numerical predictions. The models were

used to screen over 800,000 polymer–dopant combinations, recommending 500 candidates for experimentation. Guidelines highlighting the critical features for conductivity were also compiled to aid future design efforts.

Beyond electronics, a vast amount of data have been published on a wide array of doped polymers in organic photovoltaics¹⁹², including their open-circuit voltage, power conversion efficiency and other relevant parameters. Such data must be collected and curated (preferably aided by the natural language processing techniques discussed below) before they are ready for informatics, and these efforts are ongoing.

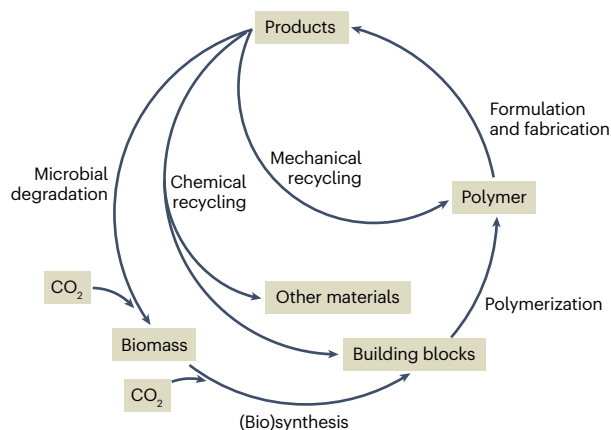
Polymers for a sustainable world

We now turn to the global issue of plastic pollution^{12,193}. An enormous volume of plastics is produced every year, but their high chemical and thermal stability makes them extremely difficult to recycle^{194–197}. According to a recent report by Greenpeace, only 5% of about 51 million tonnes of plastic created in the United States alone in 2021 was recycled¹⁹⁸, leaving the remaining for landfill at their end of life. Multiple approaches are expected to address this critical problem^{199–201}. Technical solutions, such as developing and recycling polymers, are particularly useful and active^{202–204}. Among the many classes of recyclable polymers (Fig. 7a), we address those that are biologically and chemically recyclable.

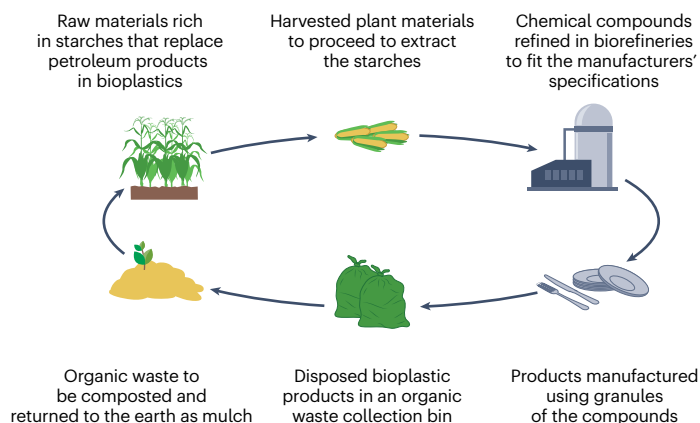
Biodegradable polymers. In biological recycling, biodegradable polymers are transformed into natural by-products such as water and CO₂ through the actions of enzymes from microorganisms such as bacteria (Fig. 7b). Several biodegradable polymer classes, including those derived from chitosan, alginate, collagen, gelatin, cellulose, hyaluronate, silk, fibrinogen and starch, have been actively considered²⁰⁵ to replace petroleum-based plastics. Applications for these polymers are targeted in numerous fields, including biomedicine²⁰⁶, the food industry^{207,208}, packaging²⁰⁹, water purification^{210,211}, electronics²¹², the automotive industry²¹³, sustainable aviation fuel²¹⁴, cosmetic products²¹⁵, fabrics, paint additives, printing and adhesives, to name just a few.

Naturally, the most important property of biodegradable polymers is their biodegradability – that is, the ability of the materials to be decomposed by enzymes. Perhaps because the biodegradation processes are highly complex and sensitive to extrinsic (such as processing) and environmental conditions, quantifying and documenting the biodegradability in a consistent manner are non-trivial²¹⁶. Commonly used measures of biodegradability are highly diverse, including weight loss²¹⁷, total organic carbon formed²¹⁸, tensile strength, carbonyl index and molecular weight change²¹⁹, and a yes/no categorical variable²²⁰ during biodegradability testing. The lack of a robust and consistent definition of biodegradability makes it challenging to create good databases for this important property. Initial steps have been taken to predict the biodegradability of polyesters using ML methods²²⁰. Clearly, to design degradable polymers for specific applications in the future, it will be necessary to predict the biodegradability of polymers

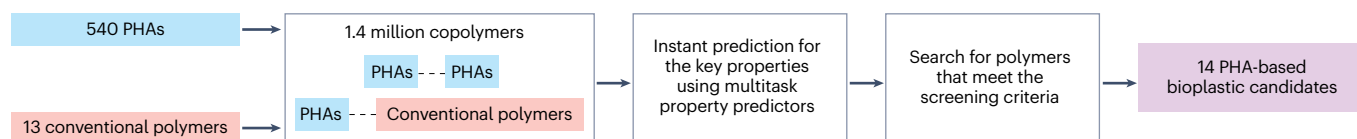
a Recyclable polymers



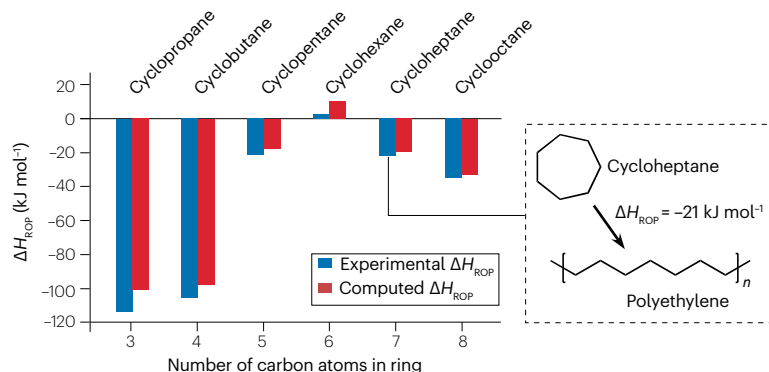
b Biorecycling: life cycle of biodegradable polymers



c Pipeline to design new PHA-based bioplastic candidates



d Ring-opening enthalpy (ΔH_{ROP}) measured and computed for cycloalkanes



e Polybenzothioane, a recyclable polymer

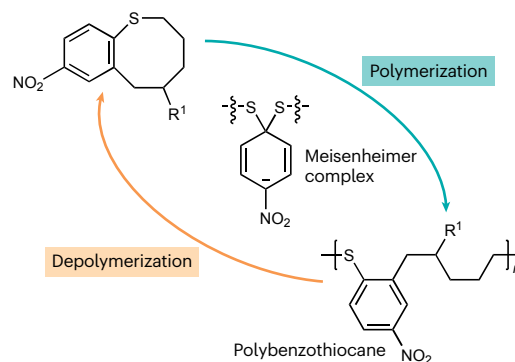


Fig. 7 | Polymers for a sustainable world. a, Regular recycling methods of recyclable polymers. **b**, The life cycle of biodegradable polymers in biorecycling. **c**, An artificial intelligence-assisted scheme to design biodegradable polymers. **d**, Measured and computed ring-opening enthalpy of cycloalkanes, showing

close correspondence between the two approaches. **e**, Polybenzothioane, a chemically recyclable polymer developed in a synergy between computational, machine learning and experimental approaches. PHA, polyhydroxyalkanoate. Panel **d** adapted with permission from ref. 68, American Chemical Society.

in a quantitative manner meeting different measures of biodegradability. This will require the development of high-quality datasets that include not only the chemical structure and biodegradability of the polymers but also other important information such as the biodegradation time profile, relevant environmental conditions, processing history, morphology and sample geometry.

Putting the quantitative predictions of biodegradability aside, and focusing only on the design of polymer chemistries amenable for biodegradation, the search space is narrowed simply to materials

occupying a known biodegradable chemical class that also display application-specific property values. Polyhydroxyalkanoates (PHAs) have emerged as a promising class of biodegradable polymers whose tremendous chemical diversity is directly accessible via biosynthesis by microalgae and bacteria. PHAs are known to be produced by about 300 species of bacteria that thrive in wastewater effluent and can be cultivated year-round, making them synthetically sustainable^{221,222}. A vast diversity of polymer compositions is possible from the over 150 PHA monomers available, and configurational

(random, alternate, block copolymer, blend and so on) and morphological (percentage crystallinity, molecular weight, dispersity and so on) degrees of freedom allow for additional tunability. However, the design rules mapping the chemistry and structure of these biopolymers onto their properties remain largely unexplored. Although DFT computations and classical MD simulations have been used for this purpose^{223–225}, predictive ML models are more reliable and scalable. In one work⁶⁷, PHA-based plastics were targeted as replacements for seven common petroleum-based consumer plastics. With this goal, multitask deep neural networks were trained on a multiproperty polymer dataset consisting of nearly 23,000 experimental data points of 13 different properties and validated on a diverse set of 15,344 homopolymers and 7,512 copolymers. Nearly 1.4 million PHA-based polymers were screened using these models, identifying two biodegradable replacements for each commodity petroleum-based plastic (Fig. 7c). Although future experimental efforts are required to validate the predictions, this work demonstrates the potential of informatics-based tools in addressing the needs of biodegradable polymer design.

Apart from the technical feasibility, it is challenging to make biopolymer production and recycling economically viable and sustainable, despite the ever-growing demand, policy-driven push, and concomitant trends in market growth for environment-friendly plastics. The cost-competitiveness and scalability of the synthetic routes for usable biopolymers, and the polymer waste recycling processes, are of prime consideration. Whereas the production cost of traditional polymers is around US\$1,000–1,500 per metric tonne, that of commonly used biopolymers can vary from 4 to 10 times more^{226,227}, owing in large part to expensive carbon substrates, the highly sterile conditions required in batch reactors during fermentation, and the laborious and time-intensive downstream processes needed to extract and purify the synthesized biopolymers²²⁸. Genetically engineering microorganisms with designer metabolic pathways that improve accumulation of biopolymer granules constitutes a future avenue worth exploring.

Chemically recyclable polymers. In chemical recycling, polymers reversibly depolymerize into monomers. Ring-opening polymers, created by opening cyclic monomers and polymerizing them, are particularly suitable because the ring-opening (that is, polymerization) and ring-closing (that is, depolymerization) reactions are easy to manipulate (Fig. 7d). On depolymerization, the monomer feedstocks can be repolymerized to create new materials with original purity and performances. The polymerization and depolymerization processes may be tuned by controllable parameters such as the monomer ring size, catalysts, temperature, solvents and other triggers^{199,200,229,230}. Research efforts in designing chemically recyclable ring-opening polymers for sustainability are timely^{69,70}.

One of the most important controllable parameters in such chemical recycling processes is ring-opening enthalpy, ΔH_{ROP} , defined as the difference between the energy of the polymer and that of the ring monomers. Ring-opening polymers that are depolymerizable should have slightly negative ΔH_{ROP} , falling roughly between -20 kJ mol^{-1} and -10 kJ mol^{-1} . Although ΔH_{ROP} can be measured experimentally, computational approaches can be much faster. This aspect is crucial for the selection of suitable monomers and eventually for the design of new depolymerizable polymers. The critical gap here is that although ΔH_{ROP} can be roughly computed in a simple and intuitive way, doing so with a satisfactory level of accuracy is non-trivial. Challenges in computing ΔH_{ROP} are diverse, including creating suitable atomic-scale polymer models, selecting the right level of theory, appropriately

sampling the polymer configurational space and, finally, reaching an ambitious level of the ‘chemical accuracy’, that is, about 5 kJ mol^{-1} or lower, expected of ab initio calculations²³¹.

A computational method has been developed⁶⁸ to quickly and accurately calculate ΔH_{ROP} for arbitrary polymers. Central to this scheme is a procedure designed to extensively sample the configuration space and compute the energies of the samples using DFT. Then, ΔH_{ROP} obtained for polymer models of different sizes is extrapolated to the limit of the polymer at infinite size. Although this method is robust and accurate, as demonstrated by the experimental and computed ring-opening enthalpies of the cycloalkane series (Fig. 7d), it is computationally demanding. To accelerate the accurate estimation of ΔH_{ROP} , a predictive ML model has been trained⁷¹ on both experimental and computed data. This model provides ΔH_{ROP} predictions with an averaged error of about 8 kJ mol^{-1} , close to the desired chemical accuracy ($\sim 5 \text{ kJ mol}^{-1}$)²³¹.

These computational and ML approaches, developed synergistically with experimental efforts, have contributed to new ring monomers that have been successfully polymerized. Synergistic computations and experiments⁶⁹ investigated a series of depolymerizable thiolactones created by systematically changing the pattern of methyl substitution and incorporation of sulfur heteroatoms. Chemically recyclable polythioethers (Fig. 7e) have also been synthesized from readily accessible benzothioecane monomers⁷⁰.

Future designs of chemically recyclable polymers will benefit from further advances in ML models. For instance, the ML approach to predict ΔH_{ROP} can be improved considerably by growing and diversifying both the computational and experimental ΔH_{ROP} datasets. VFS may be used to generate hypothetical polymers via reaction templates amenable to depolymerization. Reliable models will also need to be developed to rapidly predict other relevant properties, beyond ΔH_{ROP} , for the large number of generated polymers. Attributes of interest include thermal, mechanical, gas/solvent solubility, gas/solvent transport and thermodynamic properties, depending on the application area of interest. As with the other applications discussed above, successful materials design is an exercise in multiobjective property optimization. Beyond properties, successful and scaled-up development of chemically recyclable polymers involves synthetic considerations, such as the entropy of (de)polymerization, kinetics, solvent effects and catalyst selection¹⁹⁹.

Critical next steps

Polymer composites and formulations

In the real world, polymer composites or formulations are much more common than homogeneous neat polymers. They come in various forms, involving a base polymer matrix and additives such as reinforcing and flame-retardant materials, rheology modifiers and processing conditions. The polymer matrix is the primary continuous phase in these materials. The dispersed phase, containing the additives embedded in a discontinuous manner, and the processing protocols modulate the properties, overall appeal and utility of the polymer composites^{232,233}.

The versatility of polymer composites, allowing for desirable properties to be customized on demand, has led to their widespread applications^{234–239}. For electric vehicles, composites with suitable impact resistance, corrosion resistance, durability, and flame resistance and fire containment are used in battery enclosures. In hydrogen fuel-cell vehicles, mechanically robust composites are used to construct pressure vessels for hydrogen storage, while advancements in

polymer composites for fuel-cell membranes and ionomers are still sought after^{101–103}.

The performance of a composite is intricately linked to factors such as the chemistry and topology of the base polymer (which may also be a copolymer or polymer blend), additives and processing conditions. Although informatics protocols can handle these variations, adequate data pertaining to these variations must be captured. Progress has been made to handle copolymers and blends within informatics schemes, including representations²⁴⁰, and efforts are emerging to handle more complex composites with a variety of additives.

Traditional approaches to optimize composites, primarily involving physical manufacturing and testing, are arduous and time consuming²³⁸. ML approaches are much more efficient, either accelerating composite simulations²⁴¹ or training on experimental data to predict mechanical properties, such as fracture behaviour, ductility and density²⁴². In a recent effort, a polymer composite database was curated based on technical datasheets from major manufacturers (Tran, H. et al., unpublished work). Using handcrafted features for the polymers, fillers and processing parameters, ML models for tensile modulus and stress at break were developed (Tran, H. et al., unpublished work). A sizeable opportunity exists for developing similar models for a large spectrum of composite properties – including mechanical, thermal, flammability and transport properties – using similar approaches, a logical evolution of the AI-based protocol depicted in Box 1.

Autonomous data extraction from literature using language models

The annual growth rate of materials science papers is about 6%. Owing to the non-machine-readable nature of the content, the expanding literature makes it difficult to extract valuable quantitative and qualitative information about material properties, manually discern trends and locate materials with desirable properties. Furthermore, data corresponding to negative (or undesired) results are unlikely to be reported, skewing the balance of literature data. These factors impede progress in materials informatics, where property predictor training relies on labour-intensive data curation from literature.

Natural language processing (NLP) techniques, such as named entity recognition (NER), relation extraction, co-referencing and named entity normalization, are vital for extracting information^{243,244}. Transformer-based models like Bidirectional Encoder Representations from Transformers (BERT)^{192,245} and ChatGPT, trained on extensive unlabelled text, are predominant in self-supervised learning for contextual embeddings and information understanding. NER and relation extraction commonly use a BERT-based architecture, with labelled inputs (words and phrases indicated as material, property, characterization method and so on) feeding into task-specific ML models. Adapting these methods to new domains necessitates ontology-based labelling of new datasets. Tools like ChemDataExtractor²⁴⁶, ChemSpot²⁴⁷ and ChemicalTagger²⁴⁸ specialize in NER for material entities, but these prior NLP efforts have focused predominantly on inorganic materials^{249,250} and organic molecules^{251,252}, neglecting polymers.

Extracting information about polymers is challenging because their naming conventions vary, and these names (which in many cases do not reflect the chemical content of the polymer) cannot be directly converted to simplified molecular-input line-entry system (SMILES) strings that represent the atomic connectivity of polymers in line notation. Efforts in the past 2–3 years have taken important steps to address these polymer-specific issues^{192,243,244,251,253}. Of note is a pipeline for extracting material property data from a large body of polymer literature, derived

from a corpus of over 2.4 million materials science articles published in the past two decades¹⁹². A NER model was trained on annotated versions of the polymer text using MaterialsBERT, a language model based on PubMedBERT²⁵⁴. Using this pipeline, over 1.1 million polymer property records were extracted from the full text of the corpus. The extracted data are available at Polymer Scholar (Table 1).

Moving forward, such data and knowledge extraction endeavours must progress beyond individual sections of text, encompassing tables, figures and data dispersed throughout the full article. Each of these elements presents unique challenges. Extracting information from the full text is already intricate as relevant data may be dispersed across paragraphs, necessitating substantial advancements in relation extraction methods. Leveraging large language models such as GPT and LLaMa with tailored prompt engineering offers a promising avenue. Recognizing polymer chemical structure images and transforming them into polymer SMILES strings²⁵⁵ will also be important and may benefit from advances in molecular image recognition and SMILES conversion^{25,255}. The ultimate aspiration and imperative are to establish an autonomous pipeline commencing from published literature and patents, culminating in the extracted material property information.

Computational–experimental data fusion and multifidelity learning

In polymer informatics, an important yet underexplored opportunity lies in harnessing simulation data to construct predictive ML models. Intriguingly, the initial publicly accessible models in Polymer Genome were trained on data generated using DFT. However, challenges in producing DFT-based data for essential properties such as glass transition temperature, gas permeabilities and ionic conductivities prompted a shift towards using measured data for the model development. This approach has reached its limits, given the reliance on databases and handbooks with limited content, the struggle to capture the expanding literature data (notwithstanding the advent of NLP approaches) and the inherent limitations of physical experiments in sampling the vast chemical space.

Traditionally, the prevailing wisdom cautioned against relying solely on computer simulation data owing to certain perceived concerns. Although DFT computations are accurate, they are computationally expensive, impractical for realistic length scales and timescales, and cannot access properties of practical importance using current computational resources. Classical force fields or potentials can circumvent these challenges, but their reliability and quantitative nature are often questionable. A gap exists not only between available computational options but also between computational and experimental avenues.

Enter multifidelity or multitask learning (MTL) approaches, poised to bridge this gap. Considering two data channels, one reliable but sparse (for example, physical measurements) and the other less trustworthy but correlated to the former (for example, classical simulations of related properties), MTL approaches leverage both channels, learning correlations and predicting at the higher fidelity level while generalizing at the lower fidelity but more diverse level. Numerous success stories in this domain already exist^{34,256–258}.

The past decade witnessed the proliferation of DFT-based materials databases, such as Materials Project²⁵⁹, aflowlib²⁶⁰, OQMD²⁶¹ and NOMAD^{262,263}, but exclusively for inorganic materials. Computational data for polymers^{54,264} are limited, but we expect them to grow considerably. Extant datasets are summarized in Table 1. The anticipation for the next decade is a swift emergence of polymer databases covering DFT

and classical force-field-based data for a myriad of polymer properties, in addition to data from physical experiments; polyVERSE (briefly described in Table 1) is an initial step to alleviate this burden. Widespread availability of such databases enables computational dataset(s) to be fused with measurement datasets, to cover chemical spaces at previously unimaginable scales. MTL approaches that exploit an amalgamation of computational and experimentally measured data are expected to give rise to a new breed of predictive models.

Physics-enforced deep learning

Although data-driven approaches have substantially advanced materials discovery, the extensive reservoir of knowledge, experience, intuition, heuristics, phenomenological understanding and established relationships – collectively termed ‘known physics’ – within the domains of materials and chemistry cannot be disregarded. Integrating or imposing such known principles into ML models enhances their predictive power and ensures adherence to physical laws, leading to improved generalizability of predictions and accurate representation of physical phenomena. Furthermore, integration of known physics can address the inherent data scarcity issues prevalent in materials research.

Illustrative instances of incorporating known physics into ML models are discussed in the ‘Membranes for gas separation’ section, in which the gas permeability is elucidated as the product of gas solubility and diffusivity, and in the ‘Membranes for organic liquid mixture separations’ section, in which the organic molecule diffusivity through a polymer is conditioned on the molecular volume. The established Arrhenius relationship between ionic conductivity and temperature has also been leveraged to generate reliable predictions of lithium-ion conductivity in polymers¹⁴³, despite limited coverage in the conductivity dataset of polymer and lithium salt chemical spaces.

In practice, established equations are embedded within the loss function of neural network architectures, penalizing predictions that deviate from these principles. Physics-enforced neural networks represent a promising avenue to mitigate data scarcity, enhance generalizability and produce inherently interpretable predictive models^{265–268}.

Outlook

Through compelling use cases, we have explored the transformative impact of AI methods and informatics on accelerating polymer discovery across diverse applications, including energy-storage materials, separation membranes and sustainable materials. Several challenges must be addressed. To continuously improve intelligence, relevant high-fidelity and low-fidelity data must be captured or generated in a consistent, systematic and (re)usable manner; NLP and image analytics methods, alongside physics-driven and ML-accelerated computer simulation techniques, will be key to accumulating such data. To maximize impact, it is imperative to foster technologically relevant use cases beyond academic laboratories.

As AI-based materials intelligence ecosystems mature and evolve, they will enhance the capabilities of human researchers, driving efficiencies, accelerating discoveries and enhancing productivity. If nurtured appropriately, AI-assisted materials discovery may offer a future where the collective wisdom of intuition and experience is immortalized within the expert system, and shared and refined for the benefit of all.

References

1. Staudinger, H. Über polymerisation. *Ber. Dtsch. Chem. Ges.* **53**, 1073–1085 (1920).
2. Frey, H. & Johann, T. Celebrating 100 years of ‘polymer science’: Hermann Staudinger’s 1920 manifesto. *Polym. Chem.* **11**, 8–14 (2020).
3. Mülhaupt, R. Hermann Staudinger and the origin of macromolecular chemistry. *Angew. Chem. Int. Ed.* **43**, 1054–1063 (2004).
4. Muench, S. et al. Polymer-based organic batteries. *Chem. Rev.* **116**, 9438–9484 (2016).
5. Hubbell, J. A. Biomaterials in tissue engineering. *Nat. Biotech.* **13**, 565–576 (1995).
6. Hager, M. D., Bode, S., Weber, C. & Schubert, U. S. Shape memory polymers: past, present and future developments. *Prog. Polym. Sci.* **49–50**, 3–33 (2015).
7. Huan, T. D. et al. Advanced polymeric dielectrics for high energy density applications. *Prog. Mater. Sci.* **83**, 236–269 (2016).
8. Pan, M., Pan, C., Li, C. & Zhao, J. A review of membranes in proton exchange membrane fuel cells: transport phenomena, performance and durability. *Renew. Sust. Energ. Rev.* **141**, 110771 (2021).
9. Zhou, D., Shanmukaraj, D., Tkacheva, A., Armand, M. & Wang, G. Polymer electrolytes for lithium-based batteries: advances and prospects. *Chem* **5**, 2326–2352 (2019).
10. Baker, R. W. & Low, B. T. Gas separation membrane materials: a perspective. *Macromolecules* **47**, 6999–7013 (2014).
11. Knop, K., Hoogenboom, R., Fischer, D. & Schubert, U. S. Poly(ethylene glycol) in drug delivery: pros and cons as well as potential alternatives. *Angew. Chem. Int. Ed.* **49**, 6288–6308 (2010).
12. Geyer, R., Jambeck, J. R. & Law, K. L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **3**, e1700782 (2017).
13. Hallinan, D. T. Jr & Balsara, N. P. Polymer electrolytes. *Ann. Rev. Mater. Res.* **43**, 503–525 (2013).
14. Batra, R., Song, L. & Ramprasad, R. Emerging materials intelligence ecosystems propelled by machine learning. *Nat. Rev. Mater.* **6**, 655–678 (2021).
15. Artificial Intelligence in Chemical and Materials Science Marker report. *Future Data Stats* <https://www.futuredatastats.com/artificial-intelligence-in-chemical-and-materials-science-market> (2023).
16. Ting, J. M. & Lipscomb, C. E. Launching a materials informatics initiative for industrial applications in materials science, chemistry, and engineering. *Pure Appl. Chem.* **94**, 637 (2022).
17. Chen, L. et al. Polymer informatics: current status and critical next steps. *Mater. Sci. Eng. R Rep.* **144**, 100595 (2021).
18. Audus, D. J. & de Pablo, J. J. Polymer informatics: opportunities and challenges. *ACS Macro Lett.* **6**, 1078–1082 (2017).
19. Sha, W. et al. Machine learning in polymer informatics. *InfoMat* **3**, 353–361 (2021).
20. Mannodi-Kanakkithodi, A. et al. Scoping the polymer genome: a roadmap for rational polymer dielectrics design and beyond. *Mater. Today* **21**, 785–796 (2018).
21. Cencer, M. M., Moore, J. S. & Assary, R. S. Machine learning for polymeric materials: an introduction. *Polym. Int.* **71**, 537–542 (2022).
22. Hatakeyama-Sato, K. Recent advances and challenges in experiment-oriented polymer informatics. *Polym. J.* **55**, 117–131 (2023).
23. Fedors, R. F. A method for estimating both the solubility parameters and molar volumes of liquids. *Poly. Eng. Sci.* **14**, 147–154 (1974).
24. Kim, C., Chandrasekaran, A., Huan, T. D., Das, D. & Ramprasad, R. Polymer Genome: a data-powered polymer informatics platform for property predictions. *J. Phys. Chem. C* **122**, 17575–17585 (2018).
25. Tran, H. et al. Machine-learning predictions of polymer properties with Polymer Genome. *J. Appl. Phys.* **128**, 171104 (2020).
26. Zhao, Y., Mulder, R. J., Houshyar, S. & Le, T. C. A review on the application of molecular descriptors and machine learning in polymer design. *Polym. Chem.* **14**, 3325–3346 (2023).
27. Stuart, S., Watchorn, J. & Gu, F. X. Sizing up feature descriptors for macromolecular machine learning with polymeric biomaterials. *npj Comput. Mater.* **9**, 102 (2023).
28. Xu, C., Wang, Y. & Barati Farimani, A. TransPolymer: a transformer-based language model for polymer property predictions. *npj Comput. Mater.* **9**, 64 (2023).
29. Hatakeyama-Sato, K., Watanabe, S., Yamane, N., Igarashi, Y. & Oyaizu, K. Using GPT-4 in parameter selection of polymer informatics: improving predictive accuracy amidst data scarcity and ‘Ugly Duckling’ dilemma. *Digit. Disc.* **2**, 1548–1557 (2023).
30. Gurnani, R. et al. polyG2G: a novel machine learning algorithm applied to the generative design of polymer dielectrics. *Chem. Mater.* **33**, 7008–7016 (2021).
31. Aldeghi, M. & Coley, C. W. A graph representation of molecular ensembles for polymer property prediction. *Chem. Sci.* **13**, 10486 (2022).
32. Queen, O. et al. Polymer graph neural networks for multitask property learning. *npj Comput. Mater.* **9**, 90 (2023).
33. Kuenneth, C. & Ramprasad, R. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nat. Commun.* **14**, 4099 (2023).
34. Kuenneth, C. et al. Polymer informatics with multi-task learning. *Patterns* **2**, 100238 (2021).
35. Huan, T. D., Mannodi-Kanakkithodi, A. & Ramprasad, R. Accelerated materials property predictions and design using motif-based fingerprints. *Phys. Rev. B* **92**, 014106 (2015).
36. Mannodi-Kanakkithodi, A., Pilania, G., Huan, T. D., Lookman, T. & Ramprasad, R. Machine learning strategy for accelerated design of polymer dielectrics. *Sci. Rep.* **6**, 20952 (2016).
37. Sattari, K., Xie, Y. & Lin, J. Data-driven algorithms for inverse design of polymers. *Soft Matter* **17**, 7607–7622 (2021).
38. Afzal, M. A. F., Haghighatlari, M., Ganesh, S. P., Cheng, C. & Hachmann, J. Accelerated discovery of high-refractive-index polyimides via first-principles molecular modeling, virtual high-throughput screening, and data mining. *J. Phys. Chem. C* **123**, 14610 (2019).

Published online: 19 August 2024

39. Noh, J., Gu, G. H., Kim, S. & Jung, Y. Machine-enabled inverse design of inorganic solid materials: promises and challenges. *Chem. Sci.* **11**, 4871–4881 (2020).
40. Patra, T. K., Loeffler, T. D. & Sankaranarayanan, S. K. Accelerating copolymer inverse design using Monte Carlo tree search. *Nanoscale* **12**, 23653–23662 (2020).
41. Park, N. H. et al. A recommender system for inverse design of polycarbonates and polyesters. *Macromolecules* **53**, 10847–10854 (2020).
42. Wang, Y. et al. Toward designing highly conductive polymer electrolytes by machine learning assisted coarse-grained molecular dynamics. *Chem. Matter* **32**, 4144–4151 (2020).
43. Wu, S. et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput. Mater.* **5**, 66 (2019).
44. Khadilkar, M. R., Paradiso, S., Delaney, K. T. & Fredrickson, G. H. Inverse design of bulk morphologies in multiblock polymers using particle swarm optimization. *Macromolecules* **50**, 6702–6709 (2017).
45. Kumar, J. N. et al. Machine learning enables polymer cloud-point engineering via inverse design. *npj Comput. Mater.* **5**, 73 (2019).
46. Paradiso, S. P., Delaney, K. T. & Fredrickson, G. H. Swarm intelligence platform for multiblock polymer inverse formulation design. *ACS Macro Lett.* **5**, 972–976 (2016).
47. Coli, G. M., Boattini, E., Filion, L. & Dijkstra, M. Inverse design of soft materials via a deep learning-based evolutionary strategy. *Sci. Adv.* **8**, eabj6731 (2022).
48. Kim, C., Batra, R., Chen, L., Tran, H. & Ramprasad, R. Polymer design using genetic algorithm and machine learning. *Comput. Mater. Sci.* **186**, 110067 (2021).
49. Batra, R. et al. Polymers for extreme conditions designed using syntax-directed variational autoencoders. *Chem. Mater.* **32**, 10489–10500 (2020).
50. Jørgensen, P. B. et al. Machine learning-based screening of complex molecules for polymer solar cells. *J. Chem. Phys.* **148**, 241735 (2018).
51. Kim, S., Schroeder, C. M. & Jackson, N. E. Open macromolecular genome: generative design of synthetically accessible polymers. *ACS Polym. Au* **3**, 318–330 (2023).
52. Ohno, M., Hayashi, Y., Zhang, Q., Kaneko, Y. & Yoshida, R. SMIPoly: generation of synthesizable polymer virtual library using rule-based polymerization reactions. *J. Chem. Inf. Model.* **63**, 5539–5548 (2023).
53. Unsleber, J. P. Accelerating reaction network explorations with automated reaction template extraction and application. *J. Chem. Inf. Model.* **63**, 3392–3403 (2023).
54. Huan, T. D. et al. A polymer dataset for accelerated property prediction and design. *Sci. Data* **3**, 160012 (2016).
55. Mannodi-Kanakithodi, A. et al. Rational co-design of polymer dielectrics for energy storage. *Adv. Mater.* **28**, 6277–6291 (2016).
56. Sharma, V. et al. Rational design of all organic polymer dielectrics. *Nat. Commun.* **5**, 4845 (2014).
57. Tran, H., Shen, K. H., Shukla, S., Kwon, H. K. & Ramprasad, R. Informatics-driven selection of polymers for fuel-cell applications. *J. Phys. Chem. C* **127**, 977–986 (2023).
58. Di Noto, V., Lavina, S., Giffin, G. A., Negro, E. & Scrosati, B. Polymer electrolytes: present, past and future. *Electrochim. Acta* **57**, 4–13 (2011).
59. Meyer, W. H. Polymer electrolytes for lithium-ion batteries. *Adv. Mater.* **10**, 439–448 (1998).
60. Arya, A. & d Sharma, A. L. Polymer electrolytes for lithium ion batteries: a critical study. *Ionics* **23**, 497–540 (2017).
61. Barbosa, J. C., Gonçalves, R., Costa, C. M. & Lanceros-Méndez, S. Toward sustainable solid polymer electrolytes for lithium-ion batteries. *ASC Omega* **7**, 14457 (2022).
62. Baker, R. W. Future directions of membrane gas separation technology. *Ind. Eng. Chem. Res.* **41**, 1393–1411 (2002).
63. Marchetti, P., Jimenez Solomon, M. F., Szekely, G. & Livingston, A. G. Molecular separation with organic solvent nanofiltration: a critical review. *Chem. Rev.* **114**, 10735–10806 (2014).
64. Shi, G. M. et al. Recent progress of organic solvent nanofiltration membranes. *Prog. Polym. Sci.* **123**, 101470 (2021).
65. Semenova, S. I. Polymer membranes for hydrocarbon separation and removal. *J. Membr. Sci.* **231**, 189–207 (2004).
66. Bruno, N. C. et al. Solution-processable polytriazoles from spirocyclic monomers for membrane-based hydrocarbon separations. *Nat. Mater.* **22**, 1540–1547 (2023).
67. Kuenneth, C. et al. Bioplastic design using multitask deep neural networks. *Commun. Mater.* **3**, 96 (2022).
68. Tran, H. et al. Toward recyclable polymers: ring-opening polymerization enthalpy from first-principles. *J. Phys. Chem. Lett.* **13**, 4778–4785 (2022).
69. Stellmach, K. A. et al. Modulating polymerization thermodynamics of thiolactones through substituent and heteroatom incorporation. *ACS Macro Lett.* **11**, 895–901 (2022).
70. Su, Y. L. et al. Chemically recyclable polymer system based on nucleophilic aromatic ring-opening polymerization. *J. Am. Chem. Soc.* **145**, 13950–13956 (2023).
71. Toland, A. et al. Accelerated scheme to predict ring-opening polymerization enthalpy: simulation–experimental data fusion and multitask machine learning. *J. Phys. Chem. A* **127**, 10709–10716 (2023).
72. Feng, M. et al. Recent advances in multilayer-structure dielectrics for energy storage application. *Adv. Sci.* **8**, 2102221 (2021).
73. Feng, Q. K. et al. Recent progress and future prospects on all-organic polymer dielectrics for energy storage capacitors. *Chem. Rev.* **122**, 3820–3878 (2021).
74. Zha, J. W. & Dang, Z. M. *High Temperature Polymer Dielectrics: Fundamentals and Applications in Power Equipment* (Wiley, 2024).
75. Zha, J. W., Zheng, M. S., Fan, B. H. & Dang, Z. M. Polymer-based dielectrics with high permittivity for electric energy storage: a review. *Nano Energy* **89**, 106438 (2021).
76. Pei, J. Y. et al. All-organic dielectric polymer films exhibiting superior electric breakdown strength and discharged energy density by adjusting the electrode–dielectric interface with an organic nano-interlayer. *Energy Environ. Sci.* **14**, 5513–5522 (2021).
77. Zhu, M. X., Deng, T., Dong, L., Chen, J. M. & Dang, Z. M. Review of machine learning-driven design of polymer-based dielectrics. *IET Nanoelectr.* **5**, 24–38 (2022).
78. Luo, H. et al. Progress on polymer dielectrics for electrostatic capacitors application. *Adv. Sci.* **9**, 2202438 (2022).
79. Li, Q. et al. High-temperature dielectric materials for electrical energy storage. *Annu. Rev. Mater. Res.* **48**, 219–243 (2018).
80. Gurnani, R. et al. AI-assisted discovery of high-temperature dielectrics for energy storage. *Nat. Commun.* <https://doi.org/10.1038/s41467-024-50413-x> (2024).
81. Alamri, A. et al. Improving the rotational freedom of polyetherimide: enhancement of the dielectric properties of a commodity high-temperature polymer using a structural defect. *Chem. Mater.* **34**, 6553–6558 (2022).
82. Wu, C. et al. Rational design of all-organic flexible high-temperature polymer dielectrics. *Matter* **5**, 2615–2623 (2022).
83. Li, Z. et al. High energy density and high efficiency all-organic polymers with enhanced dipolar polarization. *J. Mater. Chem. A* **7**, 15026–15030 (2019).
84. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864 (1964).
85. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).
86. Baldwin, A. F. et al. Rational design of organotin polyesters. *Macromolecules* **48**, 2422–2428 (2015).
87. Baldwin, A. F. et al. Poly(dimethyltin glutarate) as a prospective material for high dielectric applications. *Adv. Mater.* **27**, 346–351 (2015).
88. Chen, L., Huan, T. D., Quintero, Y. C. & Ramprasad, R. Charge injection barriers at metal/polyethylene interfaces. *J. Mater. Sci.* **51**, 506–512 (2016).
89. Chen, L., Huan, T. D. & Ramprasad, R. Electronic structure of polyethylene: role of chemical, morphological and interfacial complexity. *Sci. Rep.* **7**, 6128 (2017).
90. Li, X. *Principles of Fuel Cells* (CRC, 2005).
91. Gröger, O., Gasteiger, H. A. & Suchsland, J. P. Electromobility: batteries or fuel cells? *J. Electrochem. Soc.* **162**, A2605 (2015).
92. Ogungbemi, E. et al. Fuel cell membranes — pros and cons. *Energy* **172**, 155–172 (2019).
93. Kraysberg, A. & Ein-Eli, Y. Review of advanced materials for proton exchange membrane fuel cells. *Energy Fuels* **28**, 7303–7330 (2014).
94. Kusoglu, A. & Weber, A. Z. New insights into perfluorinated sulfonic-acid ionomers. *Chem. Rev.* **117**, 987–1104 (2017).
95. Kudo, K., Jinnouchi, R. & Morimoto, Y. Humidity and temperature dependences of oxygen transport resistance of Nafion thin film on platinum electrode. *Electrochim. Acta* **209**, 682–690 (2016).
96. Kudo, K. & Morimoto, Y. Analysis of oxygen transport resistance of Nafion thin film on Pt electrode. *ECS Trans.* **50**, 1487 (2013).
97. Tanaka, T., Uchida, M. & Miyatake, K. An aromatic ionomer in the anode catalyst layer improves the start-up durability of polymer electrolyte fuel cells. *Energy Adv.* **1**, 38–44 (2022).
98. Jinnouchi, R. et al. The role of oxygen-permeable ionomer for polymer electrolyte fuel cells. *Nat. Commun.* **12**, 4956 (2021).
99. Suzuki, A. et al. Ionomer content in the catalyst layer of polymer electrolyte membrane fuel cell (PEMFC): effects on diffusion and performance. *Int. J. Hydrog. Energy* **36**, 2221–2229 (2011).
100. Jiao, K. et al. Designing the next generation of proton-exchange membrane fuel cells. *Nature* **595**, 361–369 (2021).
101. Farhat, T. R. & Hammond, P. T. Designing a new generation of proton-exchange membranes using layer-by-layer deposition of polyelectrolytes. *Adv. Funct. Mater.* **15**, 945–954 (2005).
102. Li, H. et al. Designing proton exchange membrane fuel cells with high specific power density. *J. Mater. Chem. A* **11**, 17373 (2023).
103. Zhang, G. et al. Porous flow field for next-generation proton exchange membrane fuel cells: materials, characterization, design, and challenges. *Chem. Rev.* **123**, 989–1039 (2022).
104. Miyake, J. et al. Design of flexible polyphenylene proton-conducting membrane for next-generation fuel cells. *Sci. Adv.* **3**, eaao0476 (2017).
105. Sato, K., Kajita, T. & Noro, A. Synthesis of a cross-linked polymer electrolyte membrane with an ultra-high density of sulfonic acid groups. *ACS Appl. Polym. Mater.* **5**, 3480–3488 (2023).
106. Souza, R. et al. Proton-conducting polymer electrolyte membranes based on fluoropolymers incorporating perfluorovinyl ether sulfonic acids and fluoroalkenes: synthesis and characterization. *Fuel Cell* **5**, 383–397 (2005).
107. Wang, Y. et al. Fundamentals, materials, and machine learning of polymer electrolyte membrane fuel cell technology. *Energy AI* **1**, 100014 (2020).
108. Legala, A., Zhao, J. & Li, X. Machine learning modeling for proton exchange membrane fuel cell performance. *Energy AI* **10**, 100183 (2022).
109. Mehrpooya, M., Ghorbani, B., Jafari, B., Aghbashlo, M. & Pouriman, M. Modeling of a single cell micro proton exchange membrane fuel cell by a new hybrid neural network method. *Therm. Sci. Eng. Prog.* **7**, 8–19 (2018).
110. Dekel, D. R. Review of cell performance in anion exchange membrane fuel cells. *J. Power Sources* **375**, 158–169 (2018).
111. Gottesfeld, S. et al. Anion exchange membrane fuel cells: current status and remaining challenges. *J. Power Sources* **375**, 170–184 (2018).
112. Goodenough, J. B. How we made the Li-ion rechargeable battery. *Nat. Electron.* **1**, 204 (2018).
113. Zhang, H., Zhou, M.-Y., Lin, C.-E. & Zhu, B.-K. Progress in polymeric separators for lithium ion batteries. *RSC Adv.* **5**, 89848 (2015).

114. Goriparti, S. et al. Review on recent progress of nanostructured anode materials for Li-ion batteries. *J. Power Sources* **257**, 421–443 (2014).
115. Nitta, N. & Yushin, G. High-capacity anode materials for lithium-ion batteries: choice of elements and structures for active particles. *Part. Part. Syst. Charact.* **31**, 317–336 (2014).
116. Liu, J. et al. Recent breakthroughs and perspectives of high-energy layered oxide cathode materials for lithium ion batteries. *Mater. Today* **43**, 132–165 (2021).
117. Mohamed, N. & Allam, N. K. Recent advances in the design of cathode materials for Li-ion batteries. *RSC Adv.* **10**, 21662 (2020).
118. Fergus, J. W. Recent developments in cathode materials for lithium ion batteries. *J. Power Sources* **195**, 939–954 (2010).
119. Goodenough, J. B. & Kim, Y. Challenges for rechargeable Li batteries. *Chem. Matter* **22**, 587–603 (2010).
120. Armand, M. & Tarascon, J. M. Building better batteries. *Nature* **451**, 652 (2008).
121. Tarascon, J. M. & Armand, M. Issues and challenges facing rechargeable lithium batteries. *Nature* **414**, 359–367 (2001).
122. Gupta, S., Gupta, A. K. & Pandey, B. K. First-principle study on ionic pair dissociation in PEO–PVP–NaClO₄ blend for solid polymer electrolyte. *Polym. Bull.* **79**, 4999–5018 (2022).
123. Sun, P., Bisschop, R., Niu, H. & Huang, X. A review of battery fires in electric vehicles. *Fire Technol.* **56**, 1361–1410 (2020).
124. McCune Law Group. Tesla model S and model X lithium-ion battery fires. *McCune Law Group* <https://mccunelawgroup.com/practice-areas/class-actions/tesla-model-s-and-model-x-lithium-ion-battery-fires/> (2023).
125. Irfan, U. How lithium ion batteries grounded the Dreamliner. *Sci. Am.* <https://www.scientificamerican.com/article/how-lithium-ion-batteries-grounded-the-dreamliner/> (2014).
126. Fenton, D. E., Parker, J. M. & Wright, P. V. Complexes of alkali metal ions with poly(ethylene oxide). *Polymer* **14**, 589 (1973).
127. Armand, M., Chabagno, J. M. & Duclot, M. Polymeric solid electrolytes. In *2nd International Meeting on Solid Electrolytes Abstract 6.5.1* (University of St. Andrews, 1978).
128. Webb, M. A. et al. Systematic computational and experimental investigation of lithium-ion transport mechanisms in polyester-based polymer electrolytes. *ACS Cent. Sci.* **1**, 198–205 (2015).
129. Gudla, H., Zhang, C. & Brandell, D. Effects of solvent polarity on Li-ion diffusion in polymer electrolytes: an all-atom molecular dynamics study with charge scaling. *J. Phys. Chem. B* **124**, 8124–8131 (2020).
130. Snyder, R. L. et al. Improved Li⁺ transport in polyacetal electrolytes: conductivity and current fraction in a series of polymers. *ACS Energy Lett.* **6**, 1886–1891 (2021).
131. Halat, D. M. et al. Modifying Li⁺ and anion diffusivities in polyacetal electrolytes: a pulsed-field-gradient NMR study of ion self-diffusion. *Chem. Matter* **33**, 4915–4926 (2021).
132. Sun, B., Mindemark, J., Edstrom, K. & Brandell, D. Polycarbonate-based solid polymer electrolytes for Li-ion batteries. *Solid State Ion.* **262**, 738–742 (2014).
133. Wu, I. D. & Chang, F.-C. Determination of the interaction within polyester-based solid polymer electrolyte using FTIR spectroscopy. *Polymer* **48**, 989–996 (2007).
134. Yoon, H.-K., Chung, W.-S. & Jo, N.-J. Study on ionic solid transport mechanism and interactions between salt and polymer chain in PAN based solid polymer electrolytes containing LiCF₃SO₃. *Electrochim. Acta* **50**, 289–293 (2004).
135. Kanbara, T. et al. New lithium salt ionic conductor using poly(vinyl alcohol) matrix. *Chem. Lett.* **18**, 1913–1916 (1989).
136. Tan, C. G., Siew, W. O., Pang, W. L., Osman, Z. & Chew, K. W. The effects of ceramic fillers on the PMMA-based polymer electrolyte systems. *Ionics* **13**, 361–364 (2007).
137. Bhatt, C., Swaroop, R., Arya, A. & Sharma, A. L. Effect of nano-filler on the properties of polymer nanocomposite films of PEO/PAN complexed with NaPF₆. *J. Mater. Sci. Eng. B* **5**, 418–434 (2015).
138. Rajendran, S., Sivakumar, M. & Subadevi, R. Investigations on the effect of various plasticizers in PVA–PMMA solid polymer blend electrolytes. *Mater. Lett.* **58**, 641–649 (2004).
139. Khajeh, A. et al. Early prediction of ion transport properties in solid polymer electrolytes using machine learning and system behavior-based descriptors of molecular dynamics simulations. *Macromolecules* **56**, 4787–4799 (2023).
140. Shen, X. et al. Synthesis and molecular dynamic simulation of a novel single ion conducting gel polymer electrolyte for lithium-ion batteries. *Polymer* **201**, 122568 (2020).
141. Dave, A. et al. Autonomous optimization of non-aqueous Li-ion battery electrolytes via robotic experimentation and machine learning coupling. *Nat. Commun.* **13**, 5454 (2022).
142. Wheatle, B. K., Fuentes, E. F., Lynd, N. A. & Ganesan, V. Design of polymer blend electrolytes through a machine learning approach. *Macromolecules* **53**, 9449–9459 (2020).
143. Bradford, G. et al. Chemistry-informed machine learning for polymer electrolyte discovery. *ACS Cent. Sci.* **9**, 206–216 (2023).
144. Hatakeyama-Sato, K., Uchima, Y., Kashikawa, T., Kimura, K. & Oyaizu, K. Extracting higher-conductivity designs for solid polymer electrolytes by quantum-inspired annealing. *RSC Adv.* **13**, 14651–14659 (2023).
145. High-throughput polymer design – molecular dynamics. *htpmd* <https://www.htpmd.matr.io/> (2023).
146. Hatakeyama-Sato, K., Tezuka, T., Umeki, M. & Oyaizu, K. AI-assisted exploration of superionic glass-type Li⁺ conductors with aromatic structures. *J. Am. Chem. Soc.* **142**, 3301–3305 (2020).
147. Budd, P. M. et al. Gas permeation parameters and other physicochemical properties of a polymer of intrinsic microporosity: polybenzodioxane PIM-1. *J. Membr. Sci.* **325**, 851–860 (2008).
148. Yang, Z. et al. De novo design of polymer electrolytes with high conductivity using GPT-based and diffusion-based generative models. Preprint at <https://doi.org/10.48550/arXiv.2312.06470> (2023).
149. Lei, X. et al. A self-improvable polymer discovery framework based on conditional generative model. Preprint at <https://doi.org/10.48550/arXiv.2312.04013> (2023).
150. Koros, W. J. & Zhang, C. Materials for next-generation molecularly selective synthetic membranes. *Nature* **16**, 289–297 (2017).
151. Yampolskii, Y. Polymeric gas separation membranes. *Macromolecules* **45**, 3298–3311 (2012).
152. Merkel, T. C. et al. Ultraporous, reverse-selective nanocomposite membranes. *Science* **296**, 519–522 (2002).
153. Freeman, B. D. Basis of permeability/selectivity tradeoff relations in polymeric gas separation membranes. *Macromolecules* **32**, 375–380 (1999).
154. Barnett, J. W. et al. Designing exceptional gas-separation polymer membranes using machine learning. *Sci. Adv.* **6**, eaaz4301 (2020).
155. Yang, J., Tao, L., He, J., McCutcheon, J. R. & Li, Y. Machine learning enables interpretable discovery of innovative polymers for gas separation membranes. *Sci. Adv.* **8**, eabn9545 (2022).
156. Park, H. B. et al. Polymers with cavities tuned for fast selective transport of small molecules and ions. *Science* **318**, 254–258 (2007).
157. Robeson, L. M. The upper bound revisited. *J. Membr. Sci.* **320**, 390–400 (2008).
158. Sholl, D. S. & Lively, R. P. Seven chemical separations to change the world. *Nature* **532**, 435–437 (2006).
159. Robeson, L. M. Correlation of separation factor versus permeability for polymeric membranes. *J. Membr. Sci.* **62**, 165–185 (1991).
160. Budd, P. M. Polymer with intrinsic microporosity (PIM). In *Encyclopedia of Membranes*, 1606–1607 (Springer, 2016).
161. Sridhar, S., Veerapur, R. S., Patil, M. B., Gudasi, K. B. & Aminabhavi, T. M. Matrimid polyimide membranes for the separation of carbon dioxide from methane. *J. Appl. Polym. Sci.* **106**, 1585–1594 (2007).
162. Yong, W. F., Li, F. Y., Chung, T. S. & Tong, Y. W. Highly permeable chemically modified PIM-1/matrimid membranes for green hydrogen purification. *J. Mater. Chem. A* **1**, 13914–13925 (2013).
163. Giro, R. et al. AI powered, automated discovery of polymer membranes for carbon capture. *npj Comput. Mater.* **9**, 133 (2023).
164. Gurnani, R., Kuenneth, C., Toland, A. & Ramprasad, R. Polymer informatics at scale with multitask graph neural networks. *Chem. Mater.* **35**, 1560–1567 (2023).
165. Talukder, M. J., Alshami, A. S., Tayyebi, A., Ismail, N. & Yu, X. Membrane science meets machine learning: future and potential use in assisting membrane material design and fabrication. *Sep. Purif. Rev.* <https://doi.org/10.1080/15422119.2023.2212295> (2023).
166. Yuan, Q. et al. Imputation of missing gas permeability data for polymer membranes using machine learning. *J. Membr. Sci.* **627**, 119207 (2021).
167. Park, J. Y. & Paul, D. R. Correlation and prediction of gas permeability in glassy polymer membrane materials via a modified free volume based group contribution method. *J. Membr. Sci.* **125**, 23–39 (1997).
168. Geens, J., De Witte, B. & Van der Bruggen, B. Removal of APIs (active pharmaceutical ingredients) from organic solvents by nanofiltration. *Sep. Sci. Technol.* **42**, 2435–2449 (2007).
169. Hu, J. et al. Artificial intelligence for performance prediction of organic solvent nanofiltration membranes. *J. Membr. Sci.* **619**, 118513 (2021).
170. Goebel, R. & Skibrowski, M. Machine-based learning of predictive models in organic solvent nanofiltration: pure and mixed solvent flux. *Sep. Purif. Technol.* **237**, 116363 (2020).
171. Goebel, R., Glaser, T. & Skibrowski, M. Machine-based learning of predictive models in organic solvent nanofiltration: solute rejection in pure and mixed solvents. *Sep. Purif. Technol.* **248**, 117046 (2020).
172. Ignacz, G. & Szekely, G. Deep learning meets quantitative structure–activity relationship (QSAR) for leveraging structure-based prediction of solute rejection in organic solvent nanofiltration. *J. Membr. Sci.* **646**, 120268 (2022).
173. Lee, Y. J. et al. Data-driven predictions of complex organic mixture permeation in polymer membranes. *Nat. Commun.* **14**, 4931 (2023).
174. Mathias, R. et al. Framework for predicting the fractionation of complex liquid feeds via polymer membranes. *J. Membr. Sci.* **640**, 119767 (2021).
175. Thompson, K. A. et al. N-Aryl-linked spirocyclic polymers for membrane separations of complex hydrocarbon mixtures. *Science* **369**, 310–315 (2020).
176. Inzelt, G. *Conducting Polymers: A New Era in Electrochemistry* (Springer, 2012).
177. Chiang, C. K. et al. Electrical conductivity in doped polyacetylene. *Phys. Rev. Lett.* **39**, 1098 (1977).
178. Swager, T. M. 50th anniversary perspective: conducting/semiconducting conjugated polymers. A personal perspective on the past and the future. *Macromolecules* **50**, 4867–4886 (2017).
179. Shirakawa, H. & Ikeda, S. Infrared spectra of poly(acetylene). *Polym. J.* **2**, 231–244 (1971).
180. Shirakawa, H., Louis, E. J., MacDiarmid, A. G., Chiang, C. K. & Heeger, A. J. Synthesis of electrically conducting organic polymers: halogen derivatives of polyacetylene, (CH)_x. *J. Chem. Soc. Chem. Commun.* **16**, 578–580 (1977).
181. Burroughes, J. H. et al. Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990).
182. Friend, R. H. et al. Electroluminescence in conjugated polymers. *Nature* **397**, 121–128 (1999).

183. Yang, J., Zhao, Z., Wang, S., Guo, Y. & Liu, Y. Insight into high-performance conjugated polymers for organic field-effect transistors. *Chem* **4**, 2748–2785 (2018).
184. Kim, M. et al. Donor-acceptor-conjugated polymer for high-performance organic field-effect transistors: a progress report. *Adv. Funct. Mater.* **30**, 1904545 (2020).
185. Günes, S., Neugebauer, H. & Sariciftci, N. S. Conjugated polymer-based organic solar cells. *Chem. Rev.* **107**, 1324–1338 (2007).
186. Cheng, Y. J., Yang, S. H. & Hsu, C. S. Synthesis of conjugated polymers for organic solar cell applications. *Chem. Rev.* **109**, 5868–5923 (2009).
187. Guo, B., Glavas, L. & Albertsson, A. C. Biodegradable and electrically conducting polymers for biomedical applications. *Prog. Polym. Sci.* **38**, 1263–1286 (2013).
188. Nezakati, T., Seifalian, A., Tan, A. & Seifalian, A. M. Conductive polymers: opportunities and challenges in biomedical applications. *Chem. Rev.* **118**, 6766–6843 (2018).
189. Kaur, G., Adhikari, R., Cass, P., Bown, M. & Gunatillake, P. Electrically conductive polymers and composites for biomedical applications. *RSC Adv.* **5**, 37553–37567 (2015).
190. Guimard, N. K., Gomez, N. & Schmidt, C. E. Conducting polymers in biomedical engineering. *Prog. Polym. Sci.* **32**, 876–921 (2007).
191. Sahu, H. et al. An informatics approach for designing conducting polymers. *ACS Appl. Mater. Interfaces* **13**, 53314–53322 (2021).
192. Shetty, P. et al. A general-purpose material property data extraction pipeline from large polymer corpora using natural language processing. *npj Comput. Mater.* **9**, 52 (2023).
193. Chinthapalli, R. et al. Biobased building blocks and polymers — global capacities, production and trends. *Ind. Biotechnol.* **15**, 237–241 (2018).
194. MacLeod, M., Arp, H. P. H., Tekman, M. B. & Jahnke, A. The global threat from plastic pollution. *Science* **373**, 61–65 (2021).
195. Ritchie, H. & Roser, M. Plastic pollution. *Our World in Data* <https://ourworldindata.org/plastic-pollution> (2018).
196. Thushari, G. G. N. & Senevirathna, J. D. M. Plastic pollution in the marine environment. *Heliyon* **6**, e04709 (2020).
197. Rainieri, S. & Barranco, A. Microplastics, a food safety issue? *Trends Food Sci. Technol.* **84**, 55–57 (2019).
198. Circular claims fall flat again. *Greenpeace* https://www.greenpeace.org/usa/wp-content/uploads/2022/10/GPUS_FinalReport_2022.pdf (2022).
199. Coates, G. W. & Getzler, Y. D. Chemical recycling to monomer for an ideal, circular polymer economy. *Nat. Rev. Mater.* **5**, 501–516 (2020).
200. Lange, J.-P. Sustainable development: efficiency and recycling in chemicals manufacturing. *Green Chem.* **4**, 546–550 (2002).
201. Lange, J.-P. Managing plastic waste — sorting, recycling, disposal, and product redesign. *ACS Sustain. Chem. Eng.* **9**, 15722–15738 (2021).
202. Rosenboom, J. G., Langer, R. & Traverso, G. Bioplastics for a circular economy. *Nat. Rev. Mater.* **7**, 117–137 (2022).
203. Johansen, M. R., Christensen, T. B., Ramos, T. M. & Syberg, K. A review of the plastic value chain from a circular economy perspective. *J. Environ. Manag.* **302**, 113975 (2022).
204. Kakadellis, S. & Rosetto, G. Achieving a circular bioeconomy for plastics. *Science* **373**, 49–50 (2021).
205. Nadda, A. K., Sharma, S. & Bhat, R. *Biopolymers: Recent Updates, Challenges and Opportunities* (Springer Nature, 2022).
206. Sofi, H. S. et al. Regenerated cellulose nanofibers from cellulose acetate: incorporating hydroxyapatite (HAP) and silver (Ag) nanoparticles (NPs), as a scaffold for tissue engineering applications. *Mater. Sci. Eng. C* **118**, 111547 (2021).
207. Díaz-Montes, E. & Castro-Muñoz, R. Trends in chitosan as a primary biopolymer for functional films and coatings manufacture for food and natural products. *Polymers* **13**, 767 (2021).
208. Gupta, S. et al. in *Biopolymers: Recent Updates, Challenges and Opportunities* 173–198 (Springer, 2022).
209. Ilyas, R. A. et al. Sugar palm (*Arenga pinnata* (Wurmb.) Merr) cellulosic fibre hierarchy: a comprehensive approach from macro to nano scale. *J. Mater. Res. Technol.* **8**, 2753–2766 (2019).
210. Mohamed, M. A. et al. Regenerated cellulose membrane as bio-template for in-situ growth of visible-light driven C-modified mesoporous titania. *Carbohydr. Polym.* **146**, 166–173 (2016).
211. Hernández, N., Williams, R. C. & Cochran, E. W. The battle for the ‘green’ polymer. Different approaches for biopolymer synthesis: bioadvantaged vs. bioreplacement. *Org. Biomol. Chem.* **12**, 2834–2849 (2014).
212. Ghazali, M., Triwulandari, E. & Restu, W. K. in *Biopolymers: Recent Updates, Challenges and Opportunities* 289 (Springer, 2022).
213. Ranjbar, Z., Ranjbar, B. & Foroughirad, S. in *Biopolymers: Recent Updates, Challenges and Opportunities* 271–288 (Springer, 2022).
214. Rahman, M. Z. et al. Advanced biopolymers for automobile and aviation engineering applications. *J. Polym. Res.* **30**, 106 (2023).
215. Rao, S. S., Athmika, & Rekha, P. D. in *Biopolymers: Recent Updates, Challenges and Opportunities*, 223–244 (Springer, 2022).
216. Palmisano, A. C. & Pettigrew, C. A. Biodegradability of plastics. *Bioscience* **42**, 680–685 (1992).
217. Nissa, R. C., Fikriyyah, A. K., Abdullah, A. H. D. & Pudjiraharti, S. Preliminary study of biodegradability of starch-based bioplastics using ASTM G21-70, dip-hanging, and soil burial test methods. *IOP Conf. Ser. Earth Environ. Sci.* **277**, 012007 (2019).
218. Tokiwa, Y., Calabia, B. P., Ugwu, C. U. & Aiba, S. Biodegradability of plastics. *Int. J. Mol. Sci.* **10**, 3722–3742 (2009).
219. Kong, S., Lv, X., Peng, D. & Chen, M. A new test method for biodegradability of plastics in sediment. *Environ. Technol. Innov.* **21**, 101217 (2021).
220. Fransen, K. A. et al. High-throughput experimentation for discovery of biodegradable polyesters. *Proc. Natl Acad. Sci. USA* **120**, e2220021120 (2023).
221. Kosseva, M. R. & Rusbandi, E. Trends in the biomanufacture of polyhydroxyalkanoates with focus on downstream processing. *Int. J. Biol. Macromol.* **107**, 762–778 (2018).
222. Sabbagh, F. & Muhamad, I. I. Production of poly-hydroxyalkanoate as secondary metabolite with main focus on sustainable energy. *Renew. Sust. Energ. Rev.* **72**, 95–104 (2017).
223. Bejagam, K. K., Iverson, C. N., Marrone, B. L. & Pilania, G. Molecular dynamics simulations for glass transition temperature predictions of polyhydroxyalkanoate biopolymers. *Phys. Chem. Chem. Phys.* **22**, 17880 (2020).
224. Bejagam, K. K., Iverson, C. N., Marrone, B. L. & Pilania, G. Composition and configuration dependence of glass-transition temperature in binary copolymers and blends of polyhydroxyalkanoate biopolymers. *Macromolecules* **54**, 5618–5628 (2021).
225. Bejagam, K. K. et al. Predicting the mechanical response of polyhydroxyalkanoate biopolymers using molecular dynamics simulations. *Polymers* **14**, 345 (2022).
226. Alias, N. H. et al. in *Biopolymers: Recent Updates, Challenges and Opportunities* 373–389 (Springer, 2022).
227. Chavez, B. A., Raghavan, V. & Tartakovsky, B. A comparative analysis of biopolymer production by microbial and bioelectrochemical technologies. *RSC Adv.* **12**, 16105–16118 (2022).
228. Ortelli, S. et al. in *Factories of the Future: The Italian Flagship Initiative* 131–148 (Springer, 2019).
229. Olsén, P., Odelius, K. & Albertsson, A.-C. Thermodynamic presynthetic considerations for ring-opening polymerization. *Biomacromolecules* **17**, 699–709 (2016).
230. Dubois, P., Coulembier, O. & Raquez, J.-M. *Handbook of Ring-Opening Polymerization* (Wiley Online Library, 2009).
231. Bash, P. A., Ho, L. L., MacKerell, A. D., Levine, D. & Hallstrom, P. Progress toward chemical accuracy in the computer simulation of condensed phase reactions. *Proc. Natl Acad. Sci. USA* **93**, 3698–3703 (1996).
232. Astrom, B. T. *Manufacturing of Polymer Composites* (CRC, 1997).
233. Irving, P. E. & Soutis, C. *Polymer Composites in the Aerospace Industry* (Woodhead, 2019).
234. He, X. & Wang, Y. Recent advances in the rational design of thermal conductive polymer composites. *Ind. Eng. Chem. Res.* **60**, 1137–1154 (2021).
235. Liu, G., Xiong, Y. & Zhou, L. Additive manufacturing of continuous fiber reinforced polymer composites: design opportunities and novel applications. *Compos. Commun.* **27**, 100907 (2021).
236. Zhang, F., Feng, Y. & Feng, W. Three-dimensional interconnected networks for thermally conductive polymer composites: design, preparation, properties, and mechanisms. *Mater. Sci. Eng. R Rep.* **142**, 100580 (2020).
237. Hu, J. et al. Polymer composite with improved thermal conductivity by constructing a hierarchically ordered three-dimensional interconnected network of BN. *ACS Appl. Mater. Interfaces* **9**, 13544–13553 (2017).
238. Xie, L. & Zhu, Y. Tune the phase morphology to design conductive polymer composites: a review. *Polym. Compos.* **39**, 2985–2996 (2018).
239. Cao, G., Cai, S., Zhang, H., Chen, Y. & Tian, Y. High-performance conductive polymer composites by incorporation of polyaniline-wrapped halloysite nanotubes and silver microflakes. *ACS Appl. Polym. Mater.* **4**, 3352–3360 (2022).
240. Shukla, S. S., Kuenneth, C. & Ramprasad, R. Polymer informatics beyond homopolymers. *MRS Bull.* **49**, 17–24 (2024).
241. Cassola, S., Duhovic, M., Schmidt, T. & May, D. Machine learning for polymer composites process simulation — a review. *Compos. B Eng.* **246**, 110208 (2022).
242. Sharma, A., Mukhopadhyay, T., Rangappa, S. M., Siengchin, S. & Kushvaha, V. Advances in computational intelligence of polymer composite materials: machine learning assisted modeling, analysis and design. *Arch. Comput. Methods Eng.* **29**, 3341 (2022).
243. Shetty, P. & Ramprasad, R. Machine-guided polymer knowledge extraction using natural language processing: the example of named entity normalization. *J. Chem. Inf. Model.* **61**, 5377–5385 (2021).
244. Shetty, P. & Ramprasad, R. Automated knowledge extraction from polymer literature using natural language processing. *Iscience* **24**, 101922 (2021).
245. Devlin, J., Chang, M.W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. Preprint at <https://doi.org/10.48550/arXiv.1810.04805> (2018).
246. Swain, M. C. & Cole, J. M. ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. *J. Chem. Inf. Model.* **56**, 1894–1904 (2016).
247. Rocktäschel, T., Weidlich, M. & Leser, U. ChemSpot: a hybrid system for chemical named entity recognition. *Bioinformatics* **28**, 1633–1640 (2012).
248. Hawitz, L., Jessop, D. M., Adams, N. & Murray-Rust, P. ChemicalTagger: a tool for semantic text-mining in chemistry. *J. Cheminform.* <https://doi.org/10.1186/1758-2946-3-17> (2011).
249. Wang, Z. et al. Dataset of solution-based inorganic materials synthesis procedures extracted from the scientific literature. *Sci. Data* **9**, 231 (2022).
250. Kuniyoshi, F., Ozawa, J. & Miwa, M. Analyzing research trends in inorganic materials literature using NLP. In *Proc. Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 319–334 (Springer, 2021).
251. Olivetti, E. A. et al. Data-driven materials research enabled by natural language processing and information extraction. *Appl. Phys. Rev.* **7**, 041317 (2020).
252. Öztürk, H., Özgür, A., Schwaller, P., Laino, T. & Ozkirimli, E. Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Discov. Today* **25**, 689–705 (2020).

253. Oka, H., Yoshizawa, A., Shindo, H., Matsumoto, Y. & Ishii, M. Machine extraction of polymer data from tables using XML versions of scientific articles. *Sci. Technol. Adv. Mater. Methods* **1**, 12–23 (2021).
254. Gu, Y. et al. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans. Comput. Healthc.* **3**, 1–23 (2021).
255. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **28**, 31–36 (1988).
256. Patra, A. et al. A multi-fidelity information-fusion approach to machine learn and predict polymer bandgap. *Comput. Mater. Sci.* **172**, 109286 (2020).
257. Pilania, G., Gubernatis, J. E. & Lookman, T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. *Comput. Mater. Sci.* **129**, 156–163 (2017).
258. Meng, X., Babae, H. & Karniadakis, G. E. Multi-fidelity Bayesian neural networks: algorithms and applications. *J. Comput. Phys.* **438**, 110361 (2021).
259. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
260. Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
261. 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).
262. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* **2**, 036001 (2019).
263. Draxl, C. & Scheffler, M. NOMAD: the FAIR concept for big data-driven materials science. *MRS Bull.* **43**, 676–682 (2018).
264. Hayashi, Y., Shiomi, J., Morikawa, J. & Yoshida, R. RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. *npj Comput. Mater.* **8**, 222 (2022).
265. Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
266. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
267. Ji, W., Qiu, W., Shi, Z., Pan, S. & Deng, S. Stiff-PIINN: physics-informed neural network for stiff chemical kinetics. *J. Phys. Chem. A* **125**, 8098–8106 (2021).
268. Faroughi, S. A. et al. Physics-guided, physics-informed, and physics-encoded neural networks and operators in scientific computing: fluid and solid mechanics. *J. Comput. Inf. Sci. Eng.* **24**, 040802 (2024).
269. Otsuka, S., Kuwajima, I., Hosoya, J., Xu, Y. & Yamazaki, M. PoLyInfo: Polymer database for polymeric materials design. In *2011 International Conference on Emerging Intelligent Data and Web Technologies* 22–29 (IEEE, 2011).
270. Ma, R. & Luo, T. PI1M: a benchmark database for polymer informatics. *J. Chem. Inf. Model.* **60**, 4684–4690 (2020).
271. Kamal, D. et al. Computable bulk and interfacial electronic structure features as proxies for dielectric breakdown of polymers. *ACS Appl. Mater. Interfaces* **12**, 37182 (2020).
272. Xie, T. et al. A cloud platform for sharing and automated analysis of raw data from high throughput polymer MD simulations. *APL Mach. Learn.* **1**, 046108 (2023).
273. Walsh, D. J. et al. Community Resource for Innovation in Polymer Technology (CRIPT): a scalable polymer material data structure. *ACS Cent. Sci.* **9**, 330–338 (2023).
274. Drefahl, A. CurlySMILES: a chemical language to customize and annotate encodings of molecular and nanodevice structures. *J. Cheminf.* **3**, 1 (2011).
275. Lin, T. S. et al. BigSMILES: a structurally-based line notation for describing macromolecules. *ACS Cent. Sci.* **5**, 1523–1531 (2019).
276. Schneider, L., Walsh, D., Olsen, B. & de Pablo, J. Generative BigSMILES: an extension for polymer informatics, computer simulations & ML/AI. *Digit. Discov.* **3**, 51–61 (2024).
277. Lin, T. S. et al. PolyDAT: a generic data schema for polymer characterization. *J. Chem. Inf. Model.* **61**, 1150–1163 (2021).

Acknowledgements

The authors acknowledge support from several grants from the Office of Naval Research, the National Science Foundation and Toyota Research Institute, and a grant from the Department of Energy via the Center for Understanding and Controlling Accelerated and Gradual Evolution of Materials for Energy (UNCAGE-ME), an Energy Frontier Research Center under award no. DE-SC0012577.

Author contributions

R.R. conceived and outlined the general manuscript. H.T. and R.R. wrote the initial manuscript with contributions from R.G., C.K., G.P., H.-K.K. and R.P.L. All authors edited the manuscript and figures and approved the final version for submission.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Materials* thanks the anonymous reviewers for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2024