



Megalibraries: Supercharged acceleration of materials discovery

Peter T. Smith, Carolin B. Wahl, Jenny K. Hedlund Orbeck, and Chad A. Mirkin*

Megalibraries are centimeter-scale chips containing millions to billions of individual materials each with an individually addressable position, composition, and size. Megalibraries will accelerate both the synthesis and characterization steps required for new material discoveries. Here, we review the key nanolithography developments that enable megalibrary synthesis and highlight the superior level of synthetic control within individual particles made possible by miniaturizing reactors to the nanoscale and performing chemistry confined to attoliter volumes. Massive parallelization of tip-defined materials synthesis offers control over thousands to millions of unique nanomaterials simultaneously, and then properties of interest can be rapidly screened. Unexplored materials can be readily surveyed and discovered using megalibraries, generating unprecedentedly large amounts of data that can be used to train machine learning algorithms. This approach is poised to revolutionize materials discovery that can be further amplified with highly autonomous design strategies.

Introduction: The materials discovery problem

Throughout history, society has depended on the discovery of new materials with a desired set of chemical or physical properties for utility. In the modern era, new materials have resulted in a swath of important structure–function relationships that have been used to advance applications in synthesis, health care, energy, and more. The basic principles behind traditional materials discovery processes are simple: there is a need for a material with a desired set of properties, a material is made and tested for this utility, and then optimized for performance. If the functionality is not sufficient, the process begins again and continues in a serial manner. Selecting materials to synthesize and explore in this way is a critical decision-making step, and although elemental characterization of the 118 elements across the periodic table has been established, there is much left to be explored when considering the enormous design space across all possible elemental compositions, structures, and size effects. The vastness of the materials design space can be understood in a simple thought experiment: if one considers the 61 metallic elements with stable isotopes in the periodic table, the combinations of

bi-elemental materials alone would result in more than 1800 possible combinations, with this value reaching >400 million when considering hepta-elemental compositions. Still, this preliminary estimate only accounts for composition and the sheer volume becomes even more daunting when considering the added variables of size and structure effects and estimates quickly reach the trillions. This unfathomably large materials design space is the basis for the materials discovery problem (**Figure 1a**).

For more efficient workflows, combinatorial approaches accelerate the materials design process through batch or parallelized synthesis and comparison of tens to thousands of samples.^{1–3} For example, the “Multi-Sample Concept” emerged in the late 1960s with advances in cosputtering binary and ternary compositional thin films.⁴ Beyond vapor deposition, various drop casting, inkjet printing, and microfluidic synthesis techniques have been explored in the context of combinatorial chemistry and materials science for discovery, but even these state-of-the-art high-throughput methods suffer a lack of the requisite materials diversity, synthetic fidelity, and screening resolution within each experiment for the pace of discovery to

Peter T. Smith, Department of Chemistry, Northwestern University, Evanston, USA; International Institute for Nanotechnology, Evanston, USA; ptsmith@northwestern.edu
Carolin B. Wahl, International Institute for Nanotechnology, Evanston, USA; Department of Materials Science and Engineering, Northwestern University, Evanston, USA; carolin.wahl@u.northwestern.edu

Jenny K. Hedlund Orbeck, Department of Chemistry, Northwestern University, Evanston, USA; International Institute for Nanotechnology, Evanston, USA; jenny.orbeck@northwestern.edu

Chad A. Mirkin, Department of Chemistry, Northwestern University, Evanston, USA; International Institute for Nanotechnology, Evanston, USA; Department of Materials Science and Engineering, Northwestern University, Evanston, USA; chadnano@northwestern.edu

*Corresponding author

doi:10.1557/s43577-023-00619-z

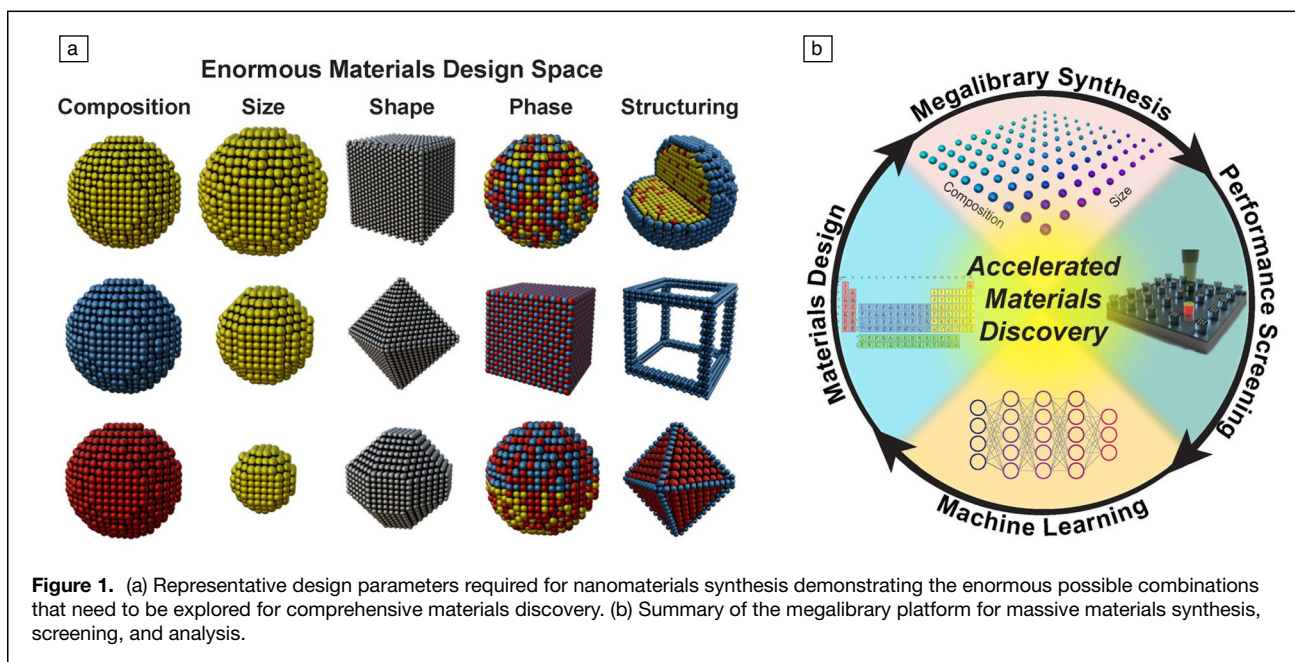


Figure 1. (a) Representative design parameters required for nanomaterials synthesis demonstrating the enormous possible combinations that need to be explored for comprehensive materials discovery. (b) Summary of the megalibrary platform for massive materials synthesis, screening, and analysis.

match modern needs.⁵ If one wants to fully interrogate the materials genome, serial and current combinatorial approaches toward materials discovery will not suffice.

In this article, we showcase the megalibrary platform as a strategy for accelerating exploration of the materials genome and assert that increasing combinatorial experimentation by multiple orders of magnitude makes megalibraries uniquely primed to revolutionize materials discovery. Megalibraries are centimeter-scale chips consisting of millions to billions of discrete and structurally characterizable materials, each of which is individually addressable. The advent of the megalibrary platform is the result of decades of research in scanning-probe lithography techniques, beginning with dippen nanolithography (DPN)⁶ and evolving to cantilever-free and massively parallel patterning⁷ with nanometer resolution on the centimeter scale.⁸ When coupled with high-throughput screening and computational approaches, such as artificial intelligence (AI), megalibraries are equipped to contend with the enormous materials design space.⁹ In summary, the megalibrary platform consisting of materials design, megalibrary synthesis, performance screening, and machine learning represents a major paradigm shift in high-throughput materials discovery that other traditional or semi-combinatorial methods cannot match (Figure 1b).

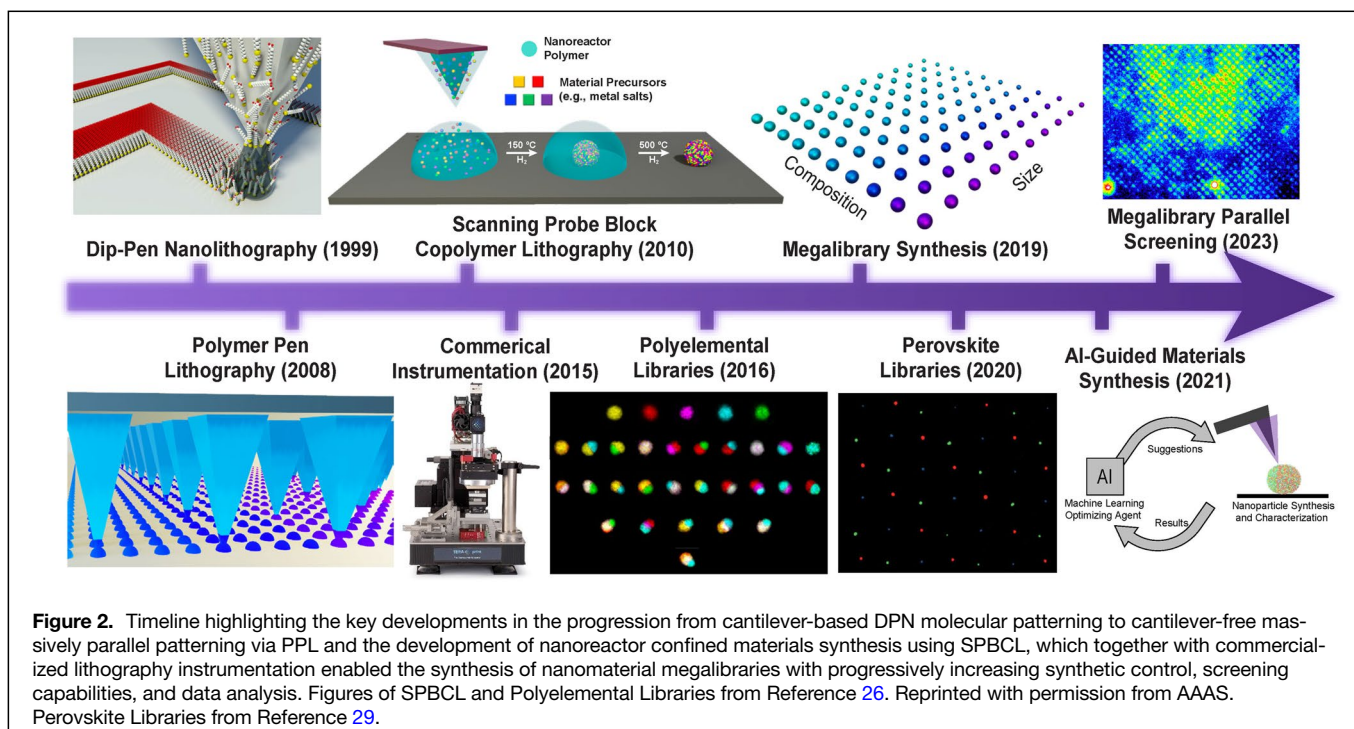
Megalibraries: Massively parallel nanomaterial synthesis

High-throughput materials discovery requires rapid synthesis and characterization tools that do not sacrifice precision or speed. Nanotechnology can solve these challenges by enabling precise materials assembly in confined spaces and providing probes with nano-to-micro-resolution analysis capabilities. Here, we review the development of nanomaterial megalibraries that contain millions of individual materials synthesized in

parallel and discuss progress made in extracting and utilizing the lessons learned from massive library screening (Figure 2).

Controlled patterning of molecules and materials with nanometer precision with DPN was introduced by the Mirkin Group in 1999.⁶ With DPN, scanning probes with nanoscale tips are coated with molecules and touched to surfaces to deposit those molecules at specific positions. DPN was a major discovery that transformed the measurement tool of atomic force microscopy (AFM) into a writing tool, enabling a broad range of new applications.¹⁰ The scale and capabilities of DPN continued to increase by introducing more scanning probe cantilevers to pattern in parallel, until 2008 yielded a cantilever-free patterning technique, termed polymer pen lithography (PPL).⁷ Massive arrays of nanoscale pen tips can be fabricated as polymer stamps using standard photolithography defined templates, and up to 11 million pen tips have been fabricated on a single centimeter-scale wafer for simultaneous patterning.¹¹

The capabilities of DPN were expanded in 2010 with scanning probe block copolymer lithography (SPBCL), transforming the writing tool into a synthesis tool.¹² In SPBCL, an aqueous ink containing poly(ethylene oxide)-*b*-poly(2-vinylpyridine) block copolymer with dissolved metal salts is patterned onto a surface using DPN to make a dome-shaped feature with <1- μm diameter of block copolymer and embedded metal salts. After thermal annealing under Ar or H₂ (~150°C), these droplets behave as “nanoreactors” as metal cations diffuse throughout the polymer matrix, are reduced, and coalesce into single nanoparticles (Figure 3a). Importantly, the surface confined nanoreactors have attoliter volumes that favor the formation of single particles within each nanoreactor to minimize surface energy. Because the nanoscale feature guides the chemistry toward a single product, the final composition is dictated by the identity and ratios



of metal precursors in the ink solution. This allows for predictive synthesis of multicomponent alloy or heterostructured nanoparticles characterized by high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) and energy-dispersive x-ray spectroscopy (EDS), whereas such stoichiometry control is difficult in traditional multicomponent nanomaterials synthesis (Figure 3b).¹³ This confined nanoreactor synthetic technique has been successfully employed for creating arrays of metal nanoparticles, metal oxides, metal sulfides, and perovskites, among others.

Beyond position and composition control (vide infra), SPBCL allows for material size and shape control (Figure 3b). Specifically, the nanoreactor volume and ink concentration dictate the precursor quantity and thus the size of the resulting single nanoparticle, first demonstrated with ~5–25-nm Au nanoparticles.¹² Extensive DPN and PPL studies have established feature size control by modulating the dwell time, force, and surface hydrophobicity, which readily transfer to SPBCL.^{14–16} Nanoparticle size can be further tuned by covalently merging the polymer nanoreactor with the metal precursors, such as poly(ethylene oxide) modified metalloporphyrins, which allowed deposition of extremely small nanoreactors that yielded sub-2-nm Pt nanoparticles.¹⁷ SPBCL typically yields spherical nanoparticles, but other shapes can be attained via post-patterning modifications. In one approach, high-index facet nanoparticles were made by alloying trace elements such as Sb, Bi, or Pb with patterned nanoparticles via solid state reaction followed by high-temperature de-alloying evaporation.^{18,19} This shape-regulating process yielded a variety of mono- or multicomponent tetrahedral

nanoparticles with simultaneous size control. Similar post-patterning modification has yielded hollow metal sulfide particles²⁰ or semiconductor nanowires.²¹

Combining these advances in SPBCL and PPL, a new modality for materials synthesis has been established termed megalibraries. Exceptional levels of nanomaterial synthetic control are merged with extremely high-throughput parallel synthesis to yield centimeter chips with millions to billions of individual particles, each with addressable position, composition, size, and shape. In 2016, PPL was improved to achieve feature separations <200 nm with a total of 5.9 billion identical features on a 14.5-cm² area.²² In a typical high-throughput megalibrary synthesis and screening experiment, 2.25-cm² chips are patterned with ~50–225 million individual nanomaterials (Figure 3c). Fabrication of such large libraries is impractical at the macroscale, but the three-dimensional (3D) miniaturization of materials synthesis with attoliter volumes and production of materials at the nanoscale provides a stand-alone solution that moves beyond conventional approaches to exploring chemical reactions. To put this in perspective, each reactor is more than a quadrillion times smaller than the 250-mL flask a traditional reaction is run in.

Next, we highlight strategies to expand control over design features, including serial composition inking, post-synthetic introduction of composition gradients, and multi-inking of pen arrays for simultaneous composition and size control. Significant strides have been made toward accessing the full capabilities of megalibrary synthesis to current capabilities that offer the synthesis and screening of >1 million unique materials per experiment.

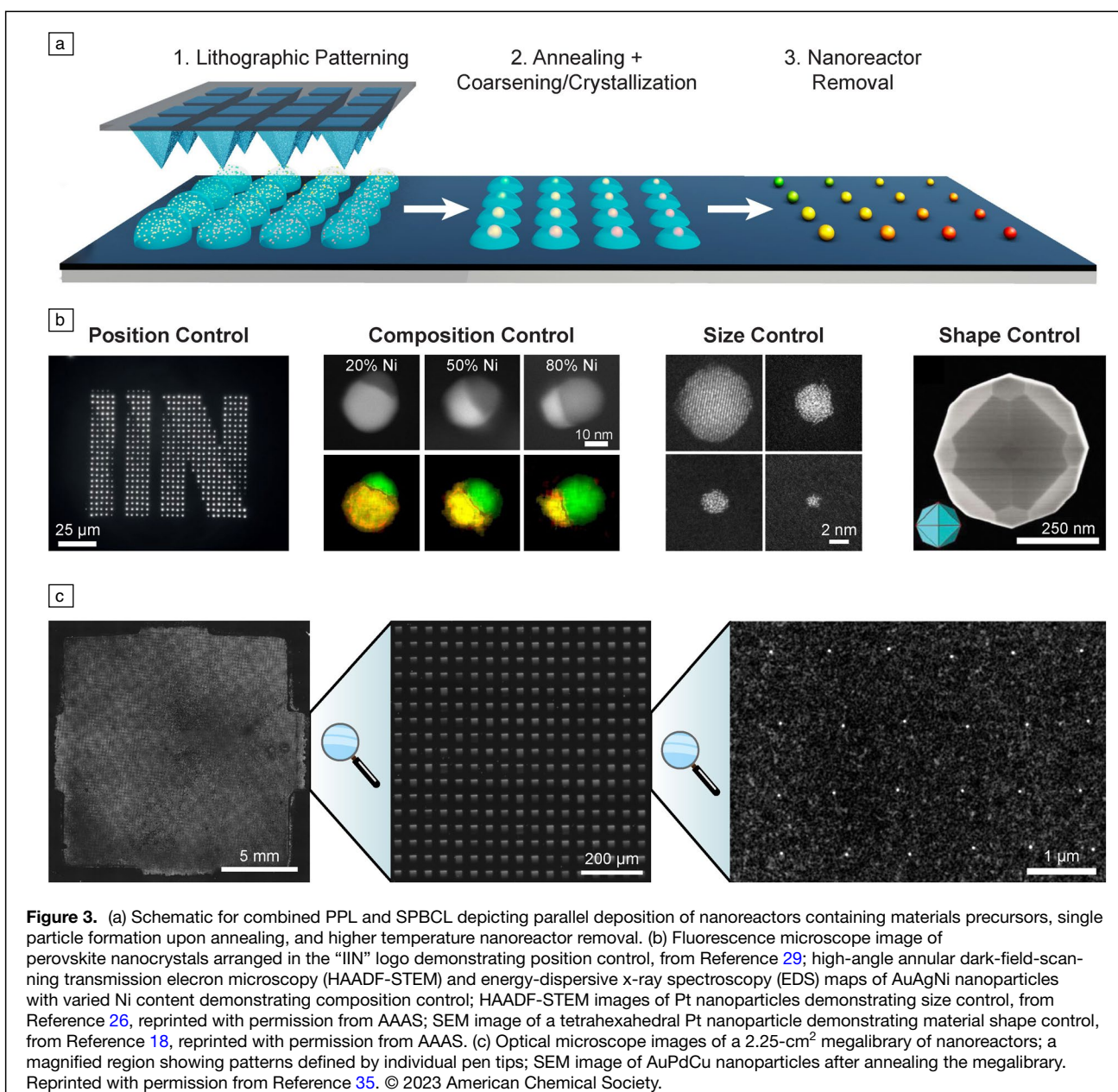


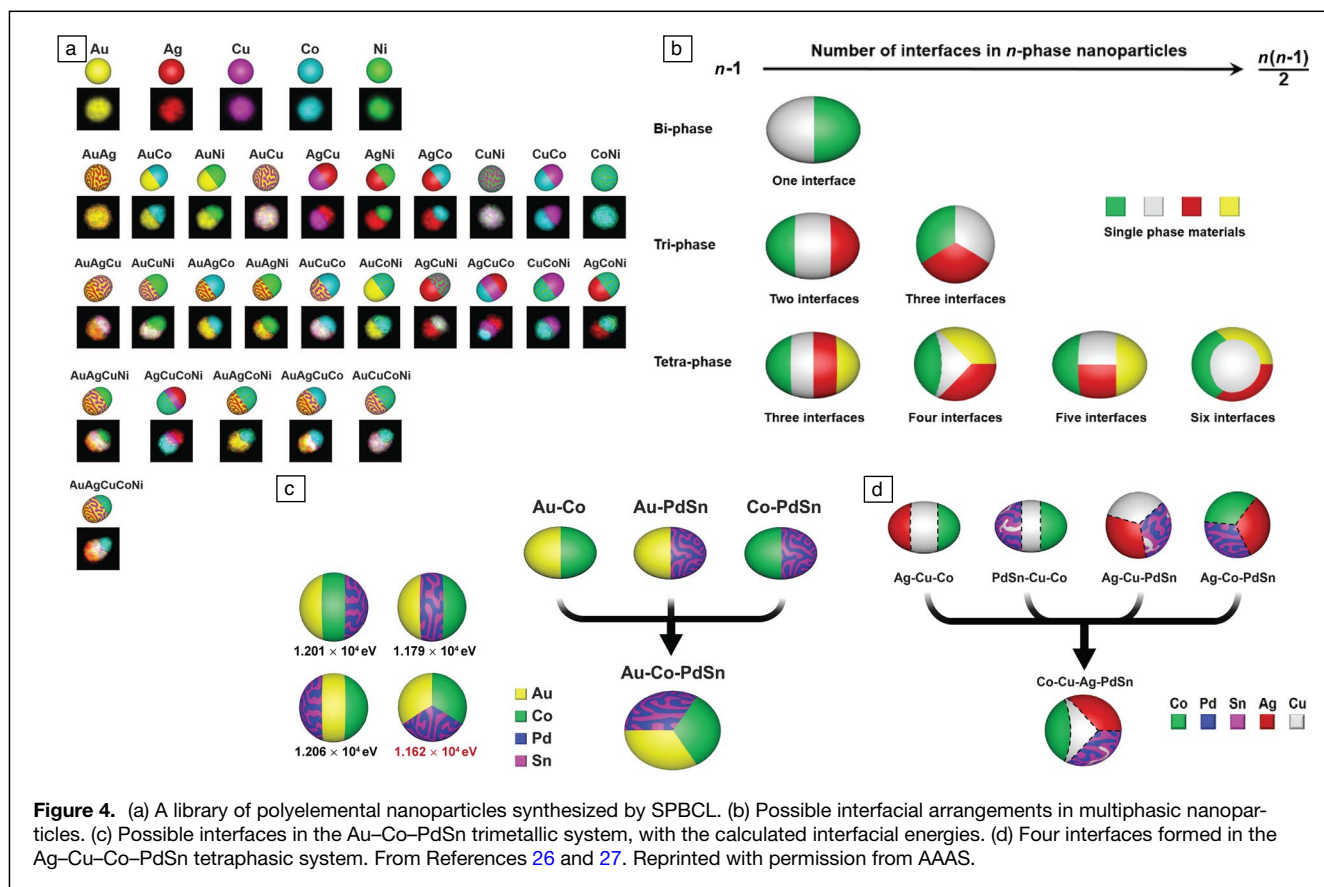
Figure 3. (a) Schematic for combined PPL and SPBCL depicting parallel deposition of nanoreactors containing materials precursors, single particle formation upon annealing, and higher temperature nanoreactor removal. (b) Fluorescence microscope image of perovskite nanocrystals arranged in the “IIN” logo demonstrating position control, from Reference 29; high-angle annular dark-field-scanning transmission electron microscopy (HAADF-STEM) and energy-dispersive x-ray spectroscopy (EDS) maps of AuAgNi nanoparticles with varied Ni content demonstrating composition control; HAADF-STEM images of Pt nanoparticles demonstrating size control, from Reference 26, reprinted with permission from AAAS; SEM image of a tetrahedral Pt nanoparticle demonstrating material shape control, from Reference 18, reprinted with permission from AAAS. (c) Optical microscope images of a 2.25-cm² megalibrary of nanoreactors; a magnified region showing patterns defined by individual pen tips; SEM image of AuPdCu nanoparticles after annealing the megalibrary. Reprinted with permission from Reference 35. © 2023 American Chemical Society.

Polyelemental nanoparticle megalibraries

Since the most critical aspect of SPBCL is the formation of a single nanoparticle in each nanoreactor, it is not immediately obvious that the same would hold for polyelemental nanoparticles. Indeed, solution-phase methods for polyelemental nanoparticle synthesis must contend with competing nucleation processes, resulting in significant compositional spread.²³ However, the extreme confinement of the precursors inside a nanoreactor guides the formation of a single polyelemental nanoparticle with precise composition control in high yield.²⁴ This was first demonstrated with a series of bimetallic and trimetallic nanoparticles,²⁵ then a representative polyelemental library of nanoparticles composed of all possible combinations of Ag, Au, Cu, Co, and Ni was synthesized (Figure 4a).²⁶

Using synthetic conditions that yield thermodynamically stable structures, the phases found in these nanoparticles generally matched expectations from bulk phase diagrams considering metal miscibility.²⁶ However, prediction of thermodynamic structures in polyelemental nanoparticles can become intractable because the bulk phase diagrams are frequently unknown. Accordingly, the promise of polyelemental nanoparticle megalibraries provides (1) spatially encoded synthesis of designer nanostructures featuring specific phases and interfaces, and (2) systematic exploration of unknown materials spaces at scale to tackle the materials discovery problem.

Synthesizing nanomaterials with a specific set of structural motifs requires establishment of corresponding design rules, especially at increasing levels of complexity. To



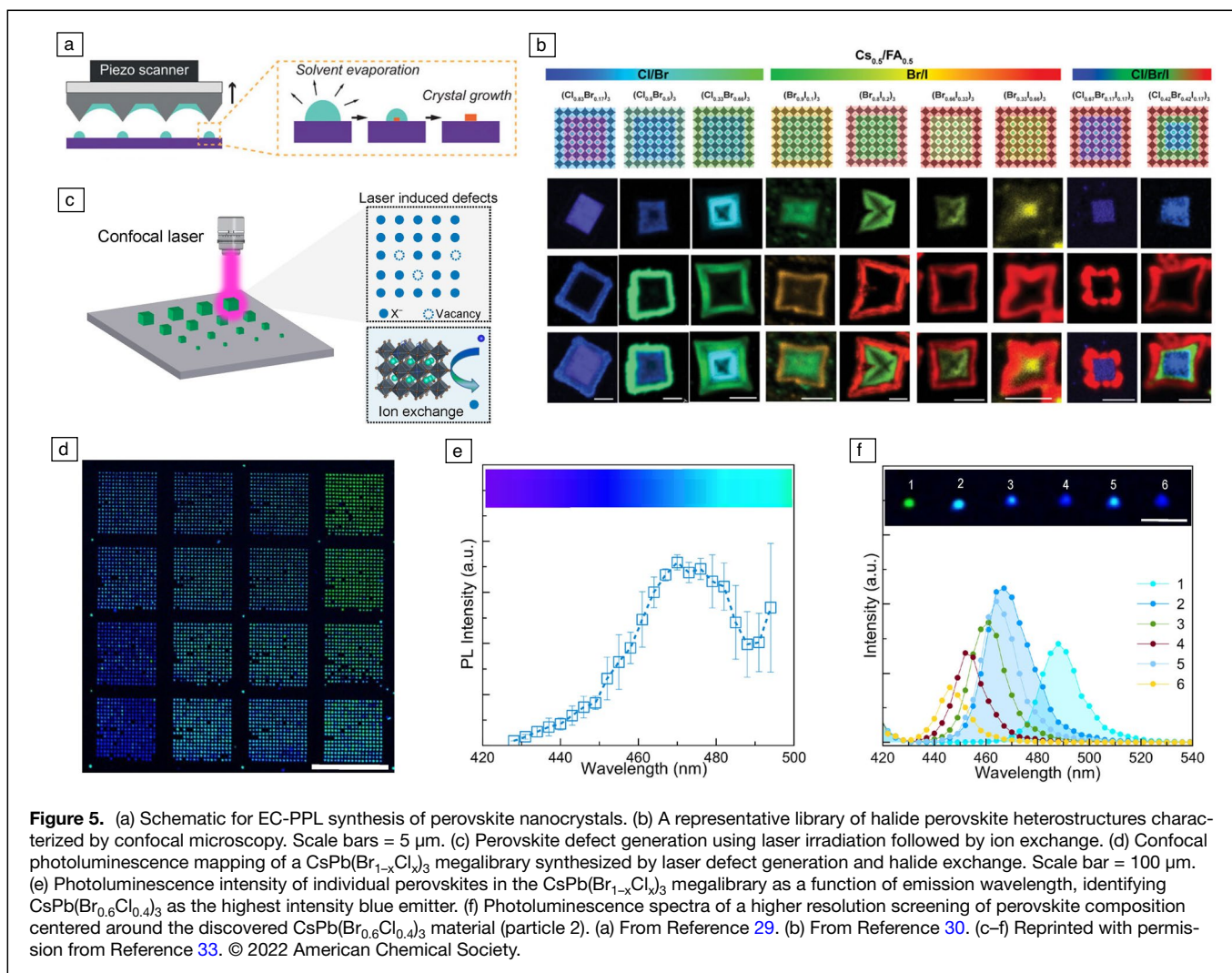
unveil such guiding principles, a systematic exploration of the Au–Ag–Cu–Co–Ni–Pd–Sn nanoparticle design space was performed, in which the elements have known but nontrivial miscibilities.²⁷ In that work, encompassing more than 30 compositions, including the seven-component single particle, nanoparticles with low phase counts (≤ 3) were sufficiently simple to allow for experimental and theoretical study, revealing two possible configurations of triphasic nanoparticles: linear (three phases with two interfaces) or pie-shaped (three phases with three interfaces) (Figure 4b). The configuration formed could not be deduced from the corresponding biphasic systems, but could be predicted theoretically as a balance of the respective surface and interfacial energies. The complexity of nanoparticles with ≥ 4 phases presented challenges for simulations; however, it was possible to predict the interfaces that formed using lessons learned from the four constituent triphasic systems (Figure 4c). Depending on their respective interface counts (two or three), the resulting tetraphasic structure had between four and six interfaces (Figure 4d).²⁷

The ability of SPBCL to reliably produce single nanoparticles of arbitrary complexity and the elucidation of the rules guiding formation of specific structural motifs enables the synthesis of designer nanostructures. Fabricated at scale, these polyelemental nanoparticle megalibraries allow for discovery of materials at an unprecedented pace.

Optoelectronics discovery using perovskite megalibraries

SPBCL has also been adapted to synthesize perovskite nanomaterials to study optical and optoelectronic properties of isolated crystals.²⁸ Through an analogous strategy, perovskite precursors are dissolved in an ink and patterned as nanoreactors, which are composed of high boiling point solvents rather than polymer. The solvent evaporates after patterning, concentrating the precursors and inducing crystallization. Combined with PPL, this evaporation-crystallization patterning, termed EC-PPL, offers excellent control over halide perovskite crystallization processes, which allows one to systematically study size/composition dependent properties and ion mixing/segregation. (Figure 5a).²⁹ For example, size-dependent emission properties of single-halide perovskite nanocrystals were first studied, revealing the importance of defects at the nanoscale. Also, perovskites compositional libraries were synthesized to make a range of heterostructures and solid solutions attainable by simply altering the precursor composition in the ink solution (Figure 5b).³⁰ The confined perovskite synthesis was critical for inducing de-mixing of ions to access 3D heterostructures difficult to obtain conventionally.

In addition to controlling perovskite composition via ink design, a post-synthetic modification strategy can transform large arrays of perovskites into megalibraries with halide composition gradients. Importantly, this enables one to address

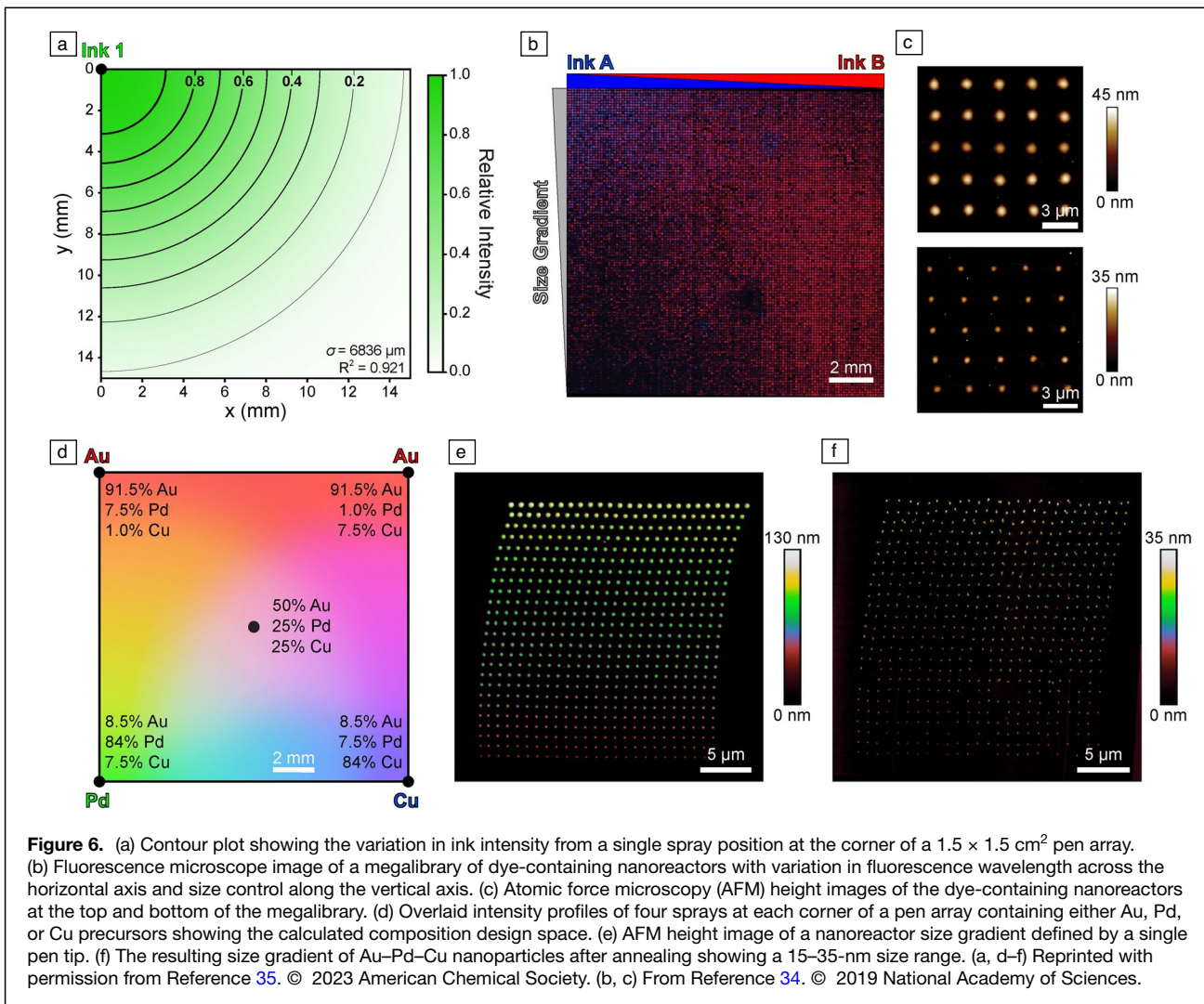


fundamental questions associated with halide perovskite structure–function relationships.²⁸ For example, defect engineering is a promising route to improve energy conversion in perovskite materials, yet progress has been slowed by few strategies to systematically vary defect concentration across large batches of samples.^{31,32} Perovskite megalibraries are a promising solution where laser exposure of individual crystals synthesized by EC-PPL induces loss of anions dependent on laser intensity and irradiation time (Figure 5c).³³ Subsequent exposure of the defective crystals to new halide solutions results in an overall halide exchange process. In a first demonstration, parallel patterning using 16 pen tips each depositing 17×17 arrays of ink yielded 4624 CsPbBr_3 perovskites with $\sim 400\text{-nm}$ diameters. Each crystal was then individually irradiated with a 405-nm laser while systematically varying the laser power and exposure time. The extent of vacancy generation was monitored by the loss in photoluminescence intensity. Finally, the particle array was submerged in a cyclohexane solution containing PbCl_2 for chloride insertion into the vacant sites and yield the combinatorial $\text{CsPb}(\text{Br}_{1-x}\text{Cl}_x)_3$ megalibrary (Figure 5d).

After perovskite megalibrary synthesis, determining the optical structure–function relationships was achieved by serial confocal mapping acquiring a photoluminescence spectrum for each particle to reveal the dependence of peak wavelength and intensity on the ratio of Cl:Br. Within the vast design space synthesized, the composition with the highest blue photoluminescence intensity was identified as $\text{CsPb}(\text{Br}_{0.6}\text{Cl}_{0.4})_3$ (Figure 5e–f). The discovery of this high-performing blue emitter was further confirmed by translation to a bulk thin-film system where this $\text{CsPb}(\text{Br}_{0.6}\text{Cl}_{0.4})_3$ composition retained its superior blue photoluminescence properties. Perovskite megalibraries are promising platforms to further push the boundaries of heterostructure synthesis and defect design at accelerated rates to uncover improved optical and optoelectronic properties.

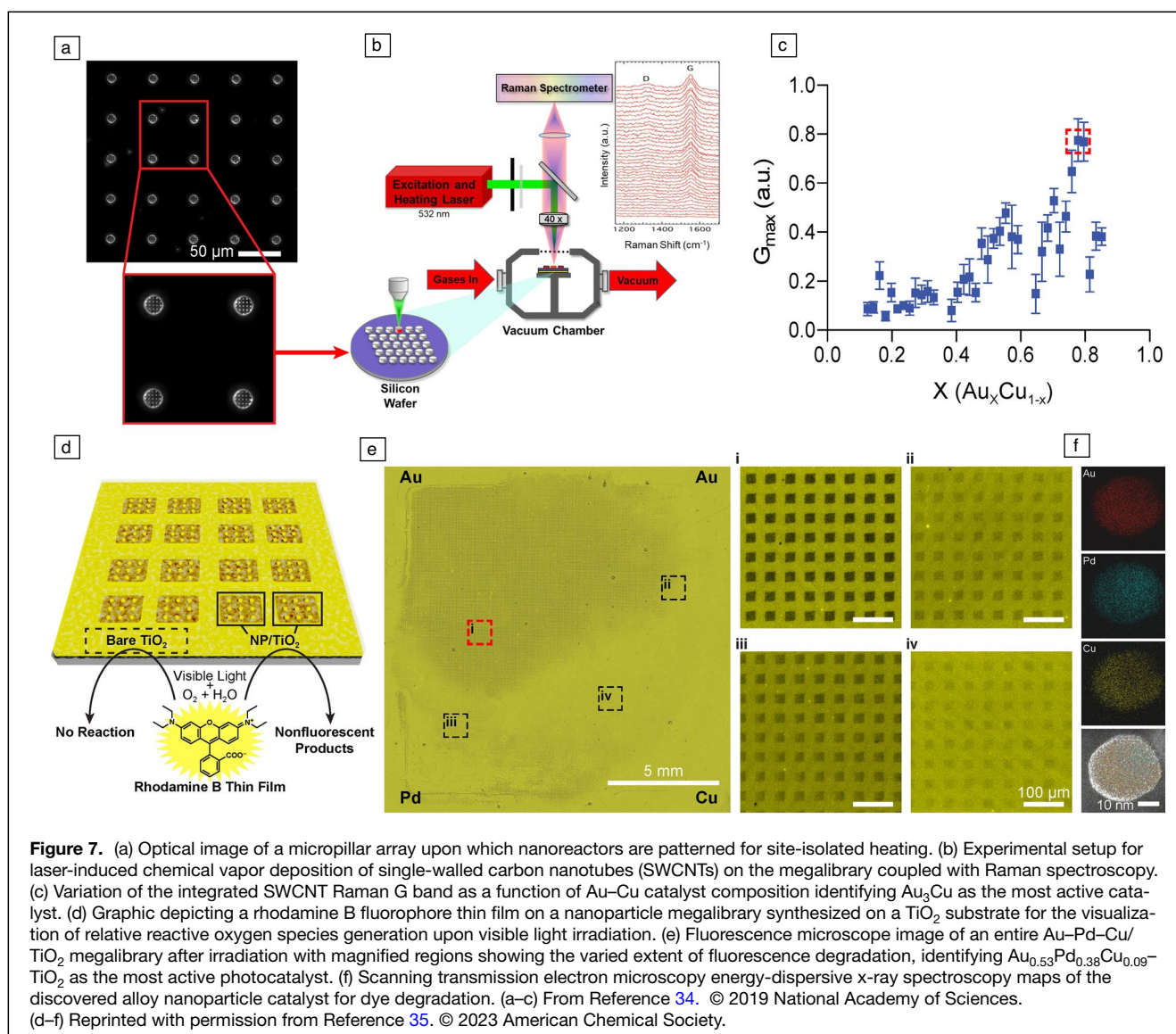
Expanding synthetic control to millions of materials

To dramatically increase synthetic control of megalibraries, it is necessary to transition away from serially altering the ink composition or serial defect generation. To accomplish this, pen arrays can be spray-coated with multiple inks containing different precursor identities/concentrations.^{34,35} Using



pressurized spray dispensers, the ink is nebulized and distributed in a radial fashion and a Gaussian distribution (Figure 6a). Overlaying multiple ink sprays onto a pen array creates composition gradients with each tip containing a unique precursor stoichiometry that is replicated in the synthesized nanomaterial after patterning and annealing. Spray inking was first demonstrated with dye nanoreactor megalibraries. Two block copolymer inks containing either the Cy5 or the rhodamine 6G fluorophores were separately sprayed at two corners of a pen array.³⁴ Resultingly, a fluorescence gradient depicted as blue-to-red was created along one axis after nanoreactor patterning (Figure 6b). Because only two sprays were performed, an ink volume gradient along the opposite array axis was also created, which translated to a nanoreactor size gradient on the substrate after patterning since more heavily inked pens deposit larger features (Figure 6c). Overall, a 126×126 pen array sprayed with two inks patterned nanoreactors with 126 colors and 126 sizes, each with 900 replicates for a total of 14,288,400 nanoreactors with 15,876 unique composition/size combinations.

Spray inking was further enhanced by spraying the four corners of a pen array to define compositional gradients along both axes and coating every pen tip with a unique precursor composition (Figure 6d). Appropriate positioning of the spray dispensers can minimize ink volume fluctuations across the array to maintain uniform nanoreactor sizes. Size gradients can be created by each individual pen tip instead of along an entire array axis by varying the pen velocity, dwell time, and force when touching the surface. This superior level of nanoparticle synthetic control was demonstrated in a three-component megalibrary covering the Au–Pd–Cu design space.³⁵ A 300×300 pen array was sprayed at the four corners, twice with Au precursors and once each with Pd or Cu precursors. Nanoreactors were deposited by each pen in 25×25 patterns where each row of nanoreactors was created using a different pen velocity and dwell time (Figure 6e). After annealing, this 2.25 cm^2 megalibrary contained a total of 56,250,000 nanoparticles with 2,250,000 unique composition/size combinations, each with 25 replicates (Figure 6f). The scale of this megalibrary and level of synthetic control attaining 1,000,000



unique nanomaterials per cm^2 far surpasses work on nanocombinatorial libraries synthesized by techniques such as drop casting, inkjet printing, or masked vapor deposition.^{1-3,36,37}

Significant SPBCL and PPL research has focused on nanomaterial structural control with patterning occurring on silicon substrates or silicon nitride membranes for electron microscopy analysis. However, such substrates limit the functionality of megalibraries because many applications rely on specific nanomaterial-substrate interactions and substrate properties, such as conductivity, light absorption, surface area, and more. Successful megalibrary synthesis requires hydrophobic surfaces, subnanometer surface roughness over micron areas, and submicron height variations over centimeter areas. Few substrates fit these requirements, but it was discovered that substrates can be temporarily modified with polystyrene thin films to enable megalibrary synthesis.³⁵ A ~ 200 -nm polystyrene thin film can be applied to an arbitrary substrate to impart the surface characteristics required for patterning and a final $>400^\circ\text{C}$

treatment induces polystyrene degradation, fully removing it and depositing the nanomaterials on the desired substrate. This was demonstrated on a range of conductors, semiconductors, and insulators and is facilitating the expansion of megalibrary screening abilities as it opens routes to explore additional material properties such as energy conversion.

Catalyst discovery using nanoparticle megalibraries

Catalyst materials are of tremendous value because most chemical processes rely on catalysis,³⁸ thus discovering materials with improved activity, selectivity, durability, and cost will always be advantageous. The first application of megalibraries for functional materials discovery was metal nanoparticle catalysts for single-walled carbon nanotube (SWCNT) synthesis demonstrated by the Mirkin Group in collaboration with the Air Force Research Laboratory.³⁴ SWCNTs are prized materials in nanotechnology with wide ranging applications necessitating improved synthesis techniques.³⁹ Spray inking

defined a Au–Cu nanoparticle design space with uniform ~2.5-nm diameters. Site-specific laser-induced heating catalyzed the vapor deposition of SWCNTs from a C₂H₄/H₂ atmosphere. To enable site-specific heating and catalysis, a specialized substrate was fabricated with 10- μ m-diameter pillars spaced 50 μ m apart (Figure 7a). PPL and thermal annealing were performed to synthesize the Au–Cu nanoparticle megalibrary on top of these pillars, and to simplify analysis, the composition gradient was divided into 36 sections. Individual pillars were heated between 700°C and 900°C and subsequently analyzed by Raman spectroscopy to quantify SWCNT formation (Figure 7b). In total, 360 SWCNT growth experiments were serially performed, and the top performing catalyst composition was determined to be Au₃Cu, a nanoparticle composition not previously known to catalyze SWCNT growth (Figure 7c).

Megalibrary synthesis using multi-spray inking and functional substrate coatings dramatically increased the number of unique materials per megalibrary. This creates challenges for screening because detection methods must have excellent spatial resolution and sensitivity to differentiate nanoparticles with micron separation. This was addressed during the discovery of photocatalytic metal nanoparticles interfaced with semiconductor materials for the visible light degradation of organic pollutants, an important strategy for environmental remediation.⁴⁰ A Au–Pd–Cu nanoparticle megalibrary was synthesized on a nanoparticulate TiO₂ substrate to endow light absorption and charge-separation properties.³⁵ Upon visible light irradiation in the presence of O₂ and H₂O, reactive oxygen species (ROS) were generated, which can degrade organic molecules. Based on this mechanism, it was reasoned that a fluorescent thin film on top of the megalibrary would lead to ROS reaction with the organic fluorophore causing a decrease in fluorescence intensity, the extent of which correlating with the amount of ROS generation. Importantly, embedding the fluorophore within a polymer film spatially confines fluorescence akin to the spatial confinement of each catalyst, thus enabling spatially selective ROS detection (Figure 7d). Photocatalyst screening was performed by irradiating the entire megalibrary for 30 min followed by fluorescence microscopy. The megalibrary contained 90,000 possible Au–Pd–Cu photocatalyst compositions, and each composition was produced in 25 \times 25 arrays with 1- μ m separation between each nanoparticle. Each array of 625 nanoparticles was separated from adjacent arrays by 25 μ m to spatially distinguish the activity of each composition. Microscopy revealed the extent of fluorescence loss surrounding each composition, revealing the relative amounts of ROS generation and thus catalyst turnovers (Figure 7e). Overlaying the fluorescence with the spray profiles identified Au_{0.53}Pd_{0.38}Cu_{0.09}–TiO₂ as the highest performing photocatalyst (Figure 7f). Parallel catalyst screening using thin-film detection is a key advance that matches the parallel materials synthesis in megalibraries and dramatically increases the speed of characterization and the amount of data generated with this single experiment screening 90,000 compositions within 1 h. Ongoing work is expanding to the single

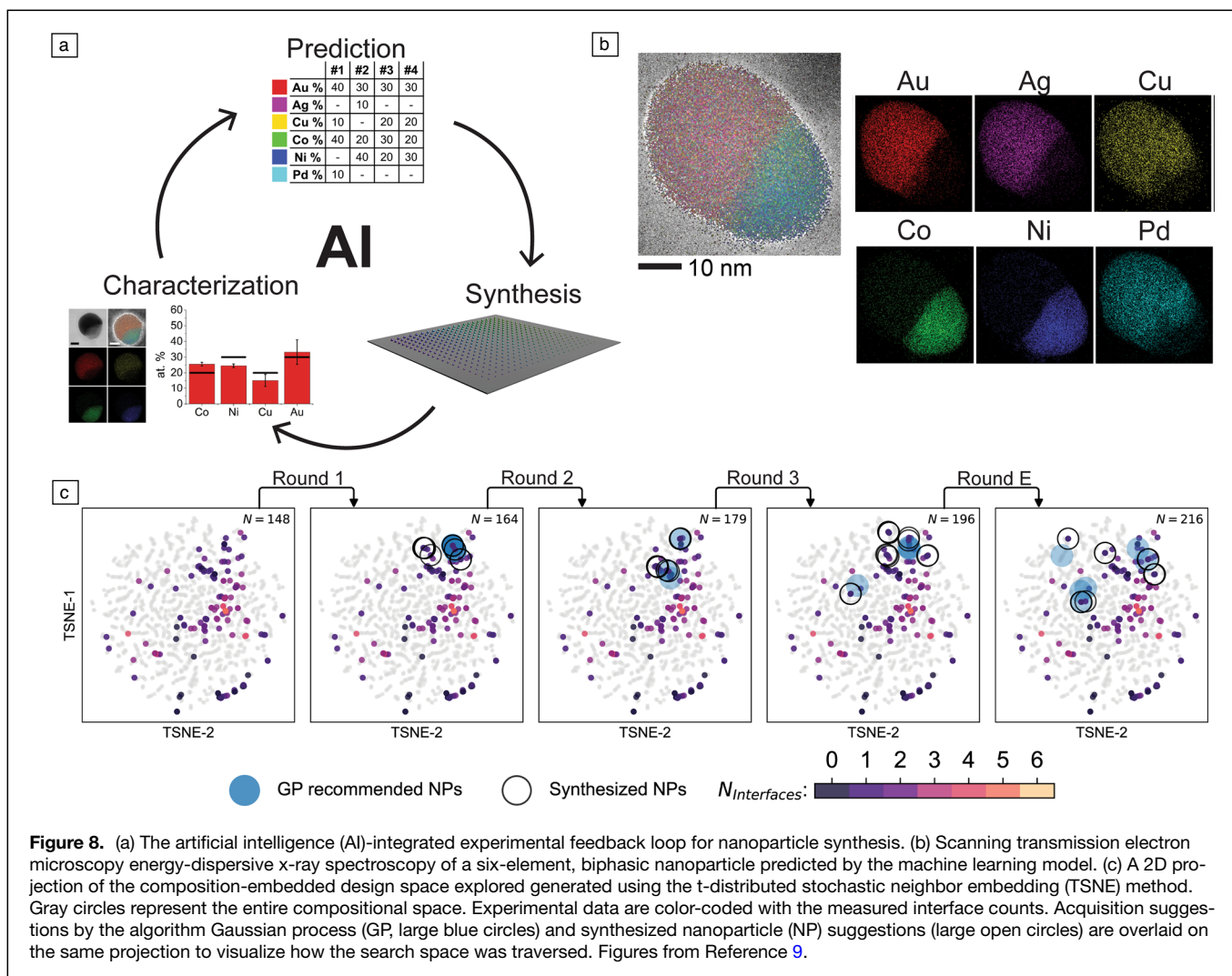
nanoparticle screening level while also broadening the immobilized fluorescence detection technique to chemoselectively react with diverse products of interest akin to intracellular small molecule sensing in chemical biology.⁴¹

Materials prediction with machine learning and AI

Given the scale of the materials discovery problem, finding new materials for specific applications generally requires *a priori* information about target design spaces and structural motifs to narrow down the otherwise intractable realm of possibilities. However, the enormous scale of this problem makes the manual compilation of this information unrealistic. The increasingly widespread adoption of AI in the physical sciences, and specifically its ability to rapidly process vast quantities of data and synthesize relevant information,⁴² suggests that the integration of AI into the megalibrary platform, and the data streams associated with it, could be an extremely powerful combination that could further accelerate materials discovery. For instance, the AlphaFold model, which predicts protein structures from their sequence, has had immense impact on the scientific community and popularized the use of predictive AI in chemistry and biology.⁴³ In the context of functional materials discovery with megalibraries, finding compositions that exhibit specific structural features is of particular interest to accelerate synthesis and screening workflows.

Since the first use of SPBCL to make multielemental nanoparticles,²⁵ it has been employed in the synthesis and structural characterization of thousands of unique compositions. These data sets serve as starting points for the guided exploration of previously unknown phase spaces. To that end, in collaboration with the Toyota Research Institute, a machine learning model was trained on the composition and interface structures of this compiled data set and used to predict specific interfacial features in an iterative feedback loop between the model and physical experiments.⁹ Specifically, the model is designed to predict elemental compositions likely to result in a desired structural feature, such as nanoparticles with a single interface. These compositions were then verified experimentally using SPBCL and STEM-EDS characterization, and the results were fed back into the model to inform the next set of experiments (Figure 8a). In the process, the model gradually discovered new composition spaces with the desired properties, predicting complex structures far beyond what would be realistic by human intuition alone. For instance, an exploration of biphasic nanoparticles produced the most chemically complex biphasic particle ever synthesized, containing six elements, Au₁₀Ag₁₀Cu₁₀Co₂₀Ni₄₀Pd₁₀ (Figure 8b). Such an iterative traversal of a given materials space occurred incrementally, venturing further away from known datapoints as more data became available (Figure 8c).⁹ Future iterative exploration of materials spaces at the megalibrary scale with predicted gradient spray profiles will enable evaluation of entire compositional spaces at once.

Beyond the predictive capabilities of AI for the targeted design and synthesis of polyelemental nanoparticles, its ability



to rapidly extract information from large multimodal data streams is perfectly suited for megalibrary screening. One can easily envision unsupervised models to cluster data generated from megalibrary chips into regions of interest. Similarly, the megalibrary geometry as regular arrays of discrete materials facilitates integration with computer vision tools for automated screening or quality control experiments. Finally, because the capabilities of AI models are derived directly from its training data sets, screening experiments with millions of datapoints are prime sources of training data with megalibraries serving as the largest available platform for continuously sourced materials data.

Summary and outlook

Megalibraries represent a transformation of high-throughput experimentation made possible with nanotechnology. Transitioning away from iterative experimentation toward parallel synthesis and screening techniques makes possible rapid surveying of unexplored materials design spaces. Experimental design does not need to be limited

by empirical observations since enormous portions of the materials genome can be readily tested, thus eliminating barriers to pursue understudied materials and applications. Importantly, the shrinking of experiments from the macro-scale to the nanoscale by performing synthetic chemistry within attoliter volume nanoreactors makes this massive parallelization possible.

From synthesis to screening, the megalibrary platform is a vastly interdisciplinary strategy. As such, this new generation of materials discovery and materials genome exploration will require contributions from materials synthesis, chemical catalysis, biological sensing, and computation, and it is thus a highly collaborative endeavor. In particular, training machine learning algorithms on the vast quantities of high-quality megalibrary data not only will help guide materials discovery but also unveil structure–function relationships that provide the basis for fundamental materials design. The combination of megalibraries and AI stands to revolutionize how we pursue and optimize new materials and bring them into real-world applications.

Acknowledgments

Research was sponsored by the Army Research Office and was accomplished under Grant No. W911NF-23-1-0141 and W911NF-23-1-0285. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. This work was also supported by the Sherman Fairchild Foundation, Inc., and the Toyota Research Institute, Inc. P.T.S. gratefully acknowledges support from the Weinberg Family Postdoctoral Fellowship.

Funding

Army Research Office, W911NF-23-1-0141, Chad A. Mirkin, W911NF-23-1-0285, Chad A. Mirkin, Toyota Research Institute, Sherman Fairchild Foundation, Weinberg College of Arts and Sciences, Northwestern University.

Data availability

Not applicable.

Conflict of interest

C.A.M. has financial interests in Mattiq, Inc., which could potentially benefit from the outcomes of this research.

References

1. E. Danielson, J.H. Golden, E.W. McFarland, C.M. Reaves, W.H. Weinberg, X.D. Wu, *Nature* **389**, 944 (1997)
2. P.J. McGinn, *ACS Comb. Sci.* **21**, 501 (2019)
3. L. Yang, J.A. Haber, Z. Armstrong, S.J. Yang, K. Kan, L. Zhou, M.H. Richter, C. Roat, N. Wagner, M. Coram, M. Berndt, P. Riley, J.M. Gregoire, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106042118 (2021)
4. J.J. Hanak, *J. Mater. Sci.* **5**, 964 (1970)
5. J.M. Gregoire, L. Zhou, J.A. Haber, *Nat. Synth.* **2**, 493 (2023)
6. R.D. Piner, J. Zhu, F. Xu, S. Hong, C.A. Mirkin, *Science* **283**, 661 (1999)
7. F. Huo, Z. Zheng, G. Zheng, L.R. Giam, H. Zhang, C.A. Mirkin, *Science* **321**, 1658 (2008)
8. K.A. Brown, J.L. Hedrick, D.J. Eichelsdoerfer, C.A. Mirkin, *ACS Nano* **13**, 8 (2019)
9. C.B. Wahl, M. Aykol, J.H. Swisher, J.H. Montoya, S.K. Suram, C.A. Mirkin, *Sci. Adv.* **7**, eabj5505 (2021)
10. G. Liu, S.H. Petrosko, Z. Zheng, C.A. Mirkin, *Chem. Rev.* **120**, 6009 (2020)
11. D.J. Eichelsdoerfer, X. Liao, M.D. Cabezas, W. Morris, B. Radha, K.A. Brown, L.R. Giam, A.B. Braunschweig, C.A. Mirkin, *Nat. Protoc.* **8**, 2548 (2013)
12. J. Chai, F. Huo, Z. Zheng, L.R. Giam, W. Shim, C.A. Mirkin, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20202 (2010)
13. J.W.M. Crawley, I.E. Gow, N. Lawes, I. Kowalec, L. Kaban, C.R.A. Catlow, A.J. Logsdail, S.H. Taylor, N.F. Dummer, G.J. Hutchings, *Chem. Rev.* **122**, 6795 (2022)
14. X. Liao, A.B. Braunschweig, Z. Zheng, C.A. Mirkin, *Small* **6**, 1082 (2010)
15. C.D. O'Connell, M.J. Higgins, D. Marusic, S.E. Moulton, G.G. Wallace, *Langmuir* **30**, 2712 (2014)
16. A. Urtizberea, M. Hirtz, H. Fuchs, *Nanofabrication* **2**, 43 (2016)
17. P.C. Chen, Y. Liu, J.S. Du, B. Meckes, V.P. Dravid, C.A. Mirkin, *J. Am. Chem. Soc.* **142**, 7350 (2020)
18. L. Huang, M. Liu, H. Lin, Y. Xu, J. Wu, V.P. Dravid, C. Wolverton, C.A. Mirkin, *Science* **365**, 1159 (2019)
19. L. Huang, H. Lin, C.Y. Zheng, E.J. Kluender, R. Golnabi, B. Shen, C.A. Mirkin, *J. Am. Chem. Soc.* **142**, 4570 (2020)

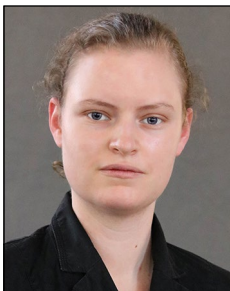
20. B. Shen, L. Huang, J. Shen, X. Hu, P. Zhong, C.Y. Zheng, C. Wolverton, C.A. Mirkin, *ACS Nano* **17**, 4642 (2023)
21. B. Shen, L. Huang, J. Shen, L. Meng, E.J. Kluender, C. Wolverton, B. Tian, C.A. Mirkin, *J. Am. Chem. Soc.* **142**, 18324 (2020)
22. J.L. Hedrick, K.A. Brown, E.J. Kluender, M.D. Cabezas, P.C. Chen, C.A. Mirkin, *ACS Nano* **10**, 3144 (2016)
23. R.E. Schaak, B.C. Steimle, J.L. Fenton, *Acc. Chem. Res.* **53**, 2558 (2020)
24. J.H. Swisher, L. Jibril, S.H. Petrosko, C.A. Mirkin, *Nat. Rev. Mater.* **7**, 428 (2022)
25. P.C. Chen, G. Liu, Y. Zhou, K.A. Brown, N. Chernyak, J.L. Hedrick, S. He, Z. Xie, Q.Y. Lin, V.P. Dravid, S.A. O'Neill-Slawecki, C.A. Mirkin, *J. Am. Chem. Soc.* **137**, 9167 (2015)
26. P.C. Chen, X. Liu, J.L. Hedrick, Z. Xie, S. Wang, Q.Y. Lin, M.C. Hersam, V.P. Dravid, C.A. Mirkin, *Science* **352**, 1565 (2016)
27. P.C. Chen, M. Liu, J.S. Du, B. Meckes, S. Wang, H. Lin, V.P. Dravid, C. Wolverton, C.A. Mirkin, *Science* **363**, 959 (2019)
28. J.S. Manser, J.A. Christians, P.V. Kamat, *Chem. Rev.* **116**, 12956 (2016)
29. J.S. Du, D. Shin, T.K. Stanev, C. Musumeci, Z. Xie, Z. Huang, M. Lai, L. Sun, W. Zhou, N.P. Stern, V.P. Dravid, C.A. Mirkin, *Sci. Adv.* **6**, eabc4959 (2020)
30. D. Shin, M. Lai, Y. Shin, J.S. Du, L. Jibril, J.M. Rondinelli, C.A. Mirkin, *Adv. Mater.* **35**, e2205923 (2023)
31. I. Ignos, S. Stavrakis, G. Nedelcu, L. Protesescu, A.J. deMello, M.V. Kovalenko, *Nano Lett.* **16**, 1869 (2016)
32. S. Chen, L. Zhang, L. Yan, X. Xiang, X. Zhao, S. Yang, B. Xu, *Adv. Funct. Mater.* **29**, 1905487 (2019)
33. M. Lai, D. Shin, L. Jibril, C.A. Mirkin, *J. Am. Chem. Soc.* **144**, 13823 (2022)
34. E.J. Kluender, J.L. Hedrick, K.A. Brown, R. Rao, B. Meckes, J.S. Du, L.M. Moreau, B. Maruyama, C.A. Mirkin, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 40 (2019)
35. P.T. Smith, Z. Ye, J. Pietryga, J. Huang, C.B. Wahl, J.K. Hedlund Orbeck, C.A. Mirkin, *J. Am. Chem. Soc.* **145**, 14031 (2023)
36. J. Wang, Y. Yoo, C. Gao, I.I. Takeuchi, X. Sun, H. Chang, X. Xiang, P.G. Schultz, *Science* **279**, 1712 (1998)
37. J.L. Hitt, Y.C. Li, S. Tao, Z. Yan, Y. Gao, S.J.L. Billinge, T.E. Mallouk, *Nat. Commun.* **12**, 1114 (2021)
38. I. Fechet, Y. Wang, J.C. Vedrine, *Catal. Today* **189**, 2 (2012)
39. M.F. De Volder, S.H. Tawfik, R.H. Baughman, A.J. Hart, *Science* **339**, 535 (2013)
40. A. Rafiq, M. Ikram, S. Ali, F. Niaz, M. Khan, Q. Khan, M. Maqbool, *J. Ind. Eng. Chem.* **97**, 111 (2021)
41. K.J. Brummer, S.W.M. Crossley, C.J. Chang, *Angew. Chem. Int. Ed.* **59**, 13734 (2020)
42. C. Suh, C. Fare, J.A. Warren, E.O. Pyzer-Knapp, *Annu. Rev. Mater. Res.* **50**, 1 (2020)
43. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohli, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zieliński, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **596**, 583 (2021) □

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.



Peter T. Smith is the Weinberg Family Postdoctoral Fellow at Northwestern University and the International Institute for Nanotechnology. He earned his BS degree from New York University and his PhD degree in chemistry from the University of California, Berkeley. His research focuses on nanoparticle megalibrary synthesis methods and solving challenges in high-throughput materials catalyst screening using molecular sensing. Smith can be reached by email at ptsmith@northwestern.edu.



Carolin B. Wahl is a doctoral candidate in materials science and engineering at Northwestern University. She received her BS degrees in chemistry and computer science from the University of California, Santa Cruz, and a master's degree in materials science and engineering from Northwestern University. Her research focuses on developing methods for rapid synthesis of nanomaterial libraries and integrating them with artificial intelligence-enhanced high-throughput structural and functional screening platforms to accelerate the discovery of new materials. Wahl can be reached by email at carolin.wahl@u.northwestern.edu.



Jenny K. Hedlund Orbeck is a scientific officer and research associate in the International Institute for Nanotechnology at Northwestern University. She earned her BS degree from the University of Iowa and PhD degree, both in chemistry, from The University of Texas at Dallas before completing a postdoctoral fellowship at the University of Wisconsin–Madison. She has also held research fellowships and appointments at Argonne National Laboratory. Her research experience spans chemistry, materials science, and nanotechnology, with an emphasis in modalities of materials characterization. Orbeck can be reached by email at jenny.orbeck@northwestern.edu.



Chad A. Mirkin is the George B. Rathmann Professor of Chemistry and founding director of the International Institute for Nanotechnology at Northwestern University. He is a chemist and nanoscience expert known for his invention of Dip-Pen Nanolithography and related nanopatterning methodologies. He is also a pioneer in nanomedicine, supramolecular chemistry, and colloidal crystal engineering with DNA. He has authored more than 850 manuscripts and >1200 patent applications (more than 400 issued worldwide), is the founder of MattIQ, Azul 3D, TeraPrint, and Flashpoint Therapeutics, and is the recipient of more than 250 national and international prizes, including the MRS Medal

(2022), Faraday Medal (UK Institution of Engineering and Technology, 2022), the Dan David Prize (2016), and King Faisal Prize in Science (Chemistry, 2023). Mirkin can be reached by email at chadnano@northwestern.edu.