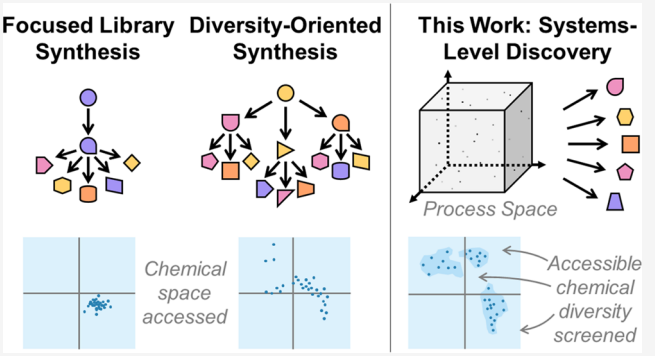
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Automated Library Generation and Serendipity Quantification Enables Diverse Discovery in Coordination Chemistry

Daniel J. Kowalski, Catriona M. MacGregor, De-Liang Long, Nicola L. Bell, and Leroy Cronin*

Cite This: J. Am. Chem. Soc. 2023, 145, 2332−2341
Read Online

ABSTRACT: Library generation experiments are a key part of the discovery of new materials, methods, and models in chemistry, but the question of how to generate high quality libraries to enable discovery is nontrivial. Herein, we use coordination chemistry to demonstrate the automation of many of the workflows used for library generation in automated hardware including the Chemputer. First, we explore the target-oriented synthesis of three influential coordination complexes, to validate key synthetic operations in our system; second, the generation of focused libraries in chemical and process space; and third, the development of a new workflow for prospecting library formation. This involved Bayesian optimization using a Gaussian process as surrogate model combined with a metric for novelty (or serendipity) quantification based on mass spectrometry data. In this way, we show directed exploration of a process space toward those areas with rarer observations and build a picture of the diversity in product distributions present across the space. We show that this effectively "engineers" serendipity into our search through the unexpected appearance of acetic anhydride, formed *in situ*, and solvent degradation products as ligands in an isolable series of three Co(III) anhydride complexes.



INTRODUCTION

Our ability to design target chemical species or synthetic routes (usually via target-oriented synthesis, TOS[1]) is predicated on previous knowledge, including models that attempt to explain patterns in chemical reactivity. New data is key to the development of materials, methods, or models, and serendipity often plays a large part in this process.[2,3] Methods that "engineer" serendipity into the chemical discovery process are of interest but hard to achieve when developing reliable approaches for library generation, for example.[4−7] Despite this, library generation is a powerful tool for gathering relevant data for pattern finding. It is possible to classify[8] libraries as either "Focused" or "Prospecting," depending on the region of chemical space accessed by members. Focused libraries are highly useful for establishing structure−activity relationships (SAR) for complex molecules; however, the diversity that can be generated through variations on the same retrosynthetic process is restricted.[8]

Prospecting libraries[8] offer an alternative focus, with an increased possibility for serendipity and diversity-oriented synthesis (DOS), which aims to develop forward synthesis methodologies toward complex molecules, is excellent at generating diverse scaffolds for medicinal applications.[1,9,10] For example, the approach of "Build/Couple/Pair" introduces diversity by mimicking the logic of biosynthetic pathways.[11−13] However, DOS is heavily reliant on multicomponent or cascade transformations, of which there are only a limited number. In addition, DOS may not always add to the understanding or development of synthetic chemistry.[10]

The development and application of automation and data science techniques have already enabled the expedience, miniaturization, and interpretation of library-generating methodologies,[14,15] but this is often focused on a limited number of well-defined processes. In this work, we apply a universal synthesis platform[16] to library generation and implement both algorithmic exploration and data science techniques to introduce a new, generally applicable workflow for prospecting library generation; see Figure 1.

We chose to apply these techniques to coordination chemistry since the highly reconfigurable and difficult-to-predict nature of these reactions,[17] particularly those in multinuclear coordination space, make this area particularly unsuitable for DOS forward synthesis methods and thus reliant on serendipity.[2,18] We address these aims by splitting the work into three phases: (i) target-oriented synthesis (TOS) of a range of coordination complexes for the validation of synthetic operations required for

Received: October 18, 2022
Published: January 17, 2023

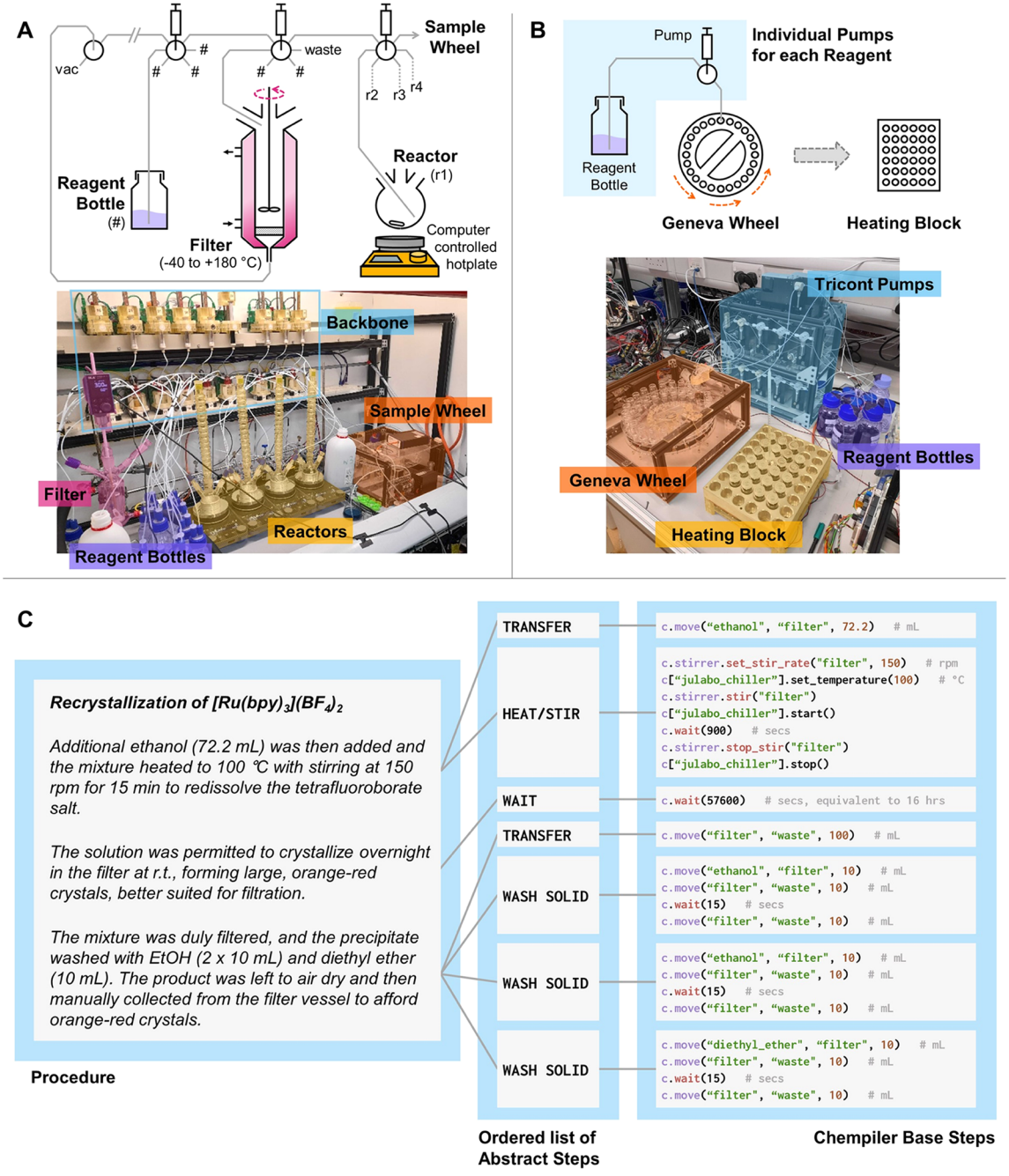


Figure 1. Chemical robots and coding logic used in this work. Simplified schematics and photographs are shown of the (A) Chemputer Platform with filter module and parallel reactors, and (B) Geneva Wheel Platform. Pump/valve combinations not accounted for in the schematic of the Chemputer are loaded with additional reagent bottles. The GWP schematic shows only a single example of the 10 pumps explicitly. (C) Overview of the structure by which automated syntheses are translated into code. Literature synthesis is broken into a series of abstract steps, to be executed in a certain order, which are then further broken into fundamental commands that can be communicated to the platform via the Chempiler package.

coordination chemistry in a universal synthesis platform, (ii) automated focused library generation, and (iii) automated generation of a prospecting library through algorithmically directed screening for chemical diversity within a process space of multinuclear species.

RESULTS AND DISCUSSION

Key Synthetic Operations for Coordination Chemistry.

The Chemputer is designed to automate liquid handling nondeterministically, meaning that chemical operations may be carried out by available modular hardware provided it is able to perform the required unit operations, regardless of topology.[19] The minimal number of modules necessary for classical coordination chemistry included here are a temperature-controlled jacketed filter module, reactor modules (i.e., RBFs), and a sample storage carousel.

This work used previously developed Chemputer modules in a topology new to synthesis (Figure 1).[20−23] Using multiple reactor modules permitted the parallelization of synthetic preparations, which could then be dispensed to vials held in the storage carousel to allow for multiple cycles of reactions before human intervention was required. Prospecting library

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341

generation used a higher-throughput setup[24] based around the Geneva wheel sample storage carousel (Figure 1). To run chemical syntheses, abstracted unit operations are used to build up a codified form of any given literature procedure as described previously.[25] The hardware is represented as a graph with nodes corresponding to individual modules (e.g., reactor, filter, input flask), and edges to the physical connections between these modules. The abstracted steps are then mapped onto the hardware to allow the same synthesis to be executed on multiple different platforms. The setup and code for all of the syntheses performed here are provided in the Supporting Information (SI), or associated GitHub repository.

Exemplar Target-Oriented Syntheses. To validate the utility of our universal synthesis platform for inorganic chemistry, three complexes were selected as targets, encompassing a range of applications, synthetic complexity, and operations common to coordination chemistry (Figure 2). $[Ru(bpy)_3](BF_4)_2$ is commonly used for photochemical applications[26−28] and was synthesized in two steps: first, the ligand substitution of ruthenium(III) chloride hydrate with 2,2′-bipyridine, via a $Ru(bpy)_2Cl_2$ intermediate, and second, the subsequent anion metathesis to afford the tetrafluoroborate salt. Precipitation yields a fine, difficult-to-handle material, which was recrystallized from ethanol to yield higher-quality single crystals. The product was afforded in a yield of 26%, in comparison to 40% manually (without recrystallization).

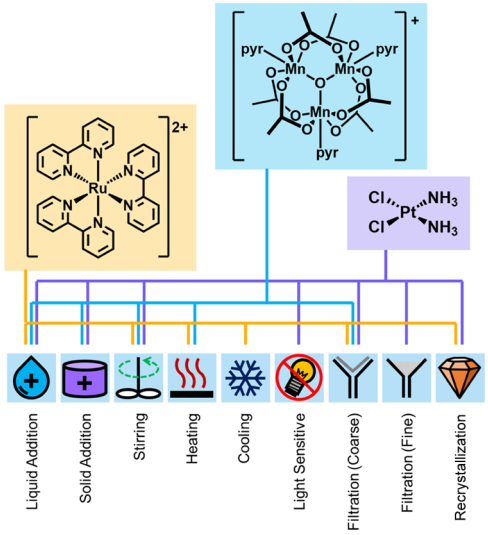


Figure 2. Synthetic operations common across coordination chemistry and their relation to the automated target-oriented syntheses conducted in this work. Coarse filtration uses the glass frit built into the jacketed filter module, and fine filtration uses a Thermo Scientific bottom-of-the-bottle filter suspended in the reactor module.

$[Mn_3O(OAc)_6(pyr)_3]ClO_4$, {$Mn_3O$}, has been widely explored in the literature as a model system in bio-inorganic chemistry,[29] as a synthetic building block for molecular materials,[30,31] and as a stoichiometric reagent.[32,33] Tetrabutylammonium permanganate, n-Bu_4NMnO_4, must be added as a solid, due to rapid degradation upon dissolution in ethanol. Two sequential solid additions were achieved through preloading the jacketed filter with n-Bu_4NMnO_4 and a reactor module with solid manganese(II) acetate. In the automated procedure, the manganese(II) acetate was first dissolved in ethanol with stirring, then combined with acetic acid and pyridine, and this mixture was transferred to the jacketed filter to initiate the conproportionation with permanganate. Given the presence of four reactor modules in addition to the jacketed filter, this process could be repeated to allow sequential addition of up to five solid reagents. The final product is precipitated via anion metathesis with $NaClO_4$ to give a yield of 49%, in comparison to 21% manually.

Cisplatin is an anticancer therapeutic and of great significance to the development of inorganic medicinal chemistry.[34,35] Our automated procedure was conducted according to a variation of Dhara's method[35,36] presented by Boreham *et al.*[37] Accessing the *cis* geometry is nontrivial, and is assured through conversion of the tetrachloroplatinate starting material to tetraiodoplatinate, where the stronger *trans*-effect of the iodo ligand promotes *cis*-substitution. The remaining iodo ligands are then removed through reaction with $AgNO_3$, and replaced with chloride. As $AgNO_3$ is photosensitive, the exclusion of light was achieved by covering the reaction vessel prior to initiating the reaction. Solid $AgNO_3$ was added to this occluded reactor, along with the platinate starting material, to remove the need for storage and transference of a light-sensitive solution through the backbone. The fine precipitate of AgI generated on reaction contaminated the product in initial runs and required the addition of a disposable in-line filter through which the reaction mixture was passed to reach the backbone. Cisplatin was collected with a yield of 30%, in comparison to 43% manually.

In addition to the coordination complexes, both metastable polymorphs of calcium carbonate—vaterite and aragonite—were prepared from a common pool of reagents.[38] This system is relatively sensitive to the synthetic environment and will readily convert to the thermodynamically stable calcite polymorph if not appropriately controlled. These syntheses speak to the fine control of temperature, stir rate, and addition rate, and to the meticulousness of filtration in the Chemputer. These aspects are likely to be advantageous in classical coordination chemistry, particularly during sensitive processes such as recrystallization.

Focused Library Generation. With the synthetic operations validated, we next explored the development of noniterative focused libraries via two methods: (i) combinatorial reaction of a series of ligands with a metal complex and (ii) screening a range of values for a key synthetic parameter.

Combinatorial production of focused molecular libraries by varying reagent combinations is key to determining structure−activity relationships.[8] For this reason, the lead-optimization phase of medicinal discovery chemistry often involves late-stage combinatorial variation of pendant structural features attached to a core scaffold.[39] As such, we automated the generation of a focused library via a combination of the $Ru(bpy)_2Cl_2$ intermediate from the synthesis of $[Ru(bpy)_3]^{2+}$ with a set of six *N*-heterocycles. Products were analyzed via MS and UV−vis spectroscopy (Figure 3, left).

$Ru(bpy)_2Cl_2$ was synthesized autonomously in the jacketed filter as previously and redissolved to serve as the metal precursor stock. The four reactors were operated in parallel, and the resultant product mixtures were transferred to the sample storage carousel. The reactor modules were then cleaned, and the next set of reactions was attempted (see the SI). A sequence of 20 reactions was shown to run consecutively without the need for human interaction, with only a single minor fault (one instance of the carousel failing to turn before dispensing). All of the heterocycles induced the formation of new complexes, except for 2,6-lutidine. 2-picolinic acid produced complexes containing a chelating κ^2-carboxylate ligand. 4,4′-Bipyridine,

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341

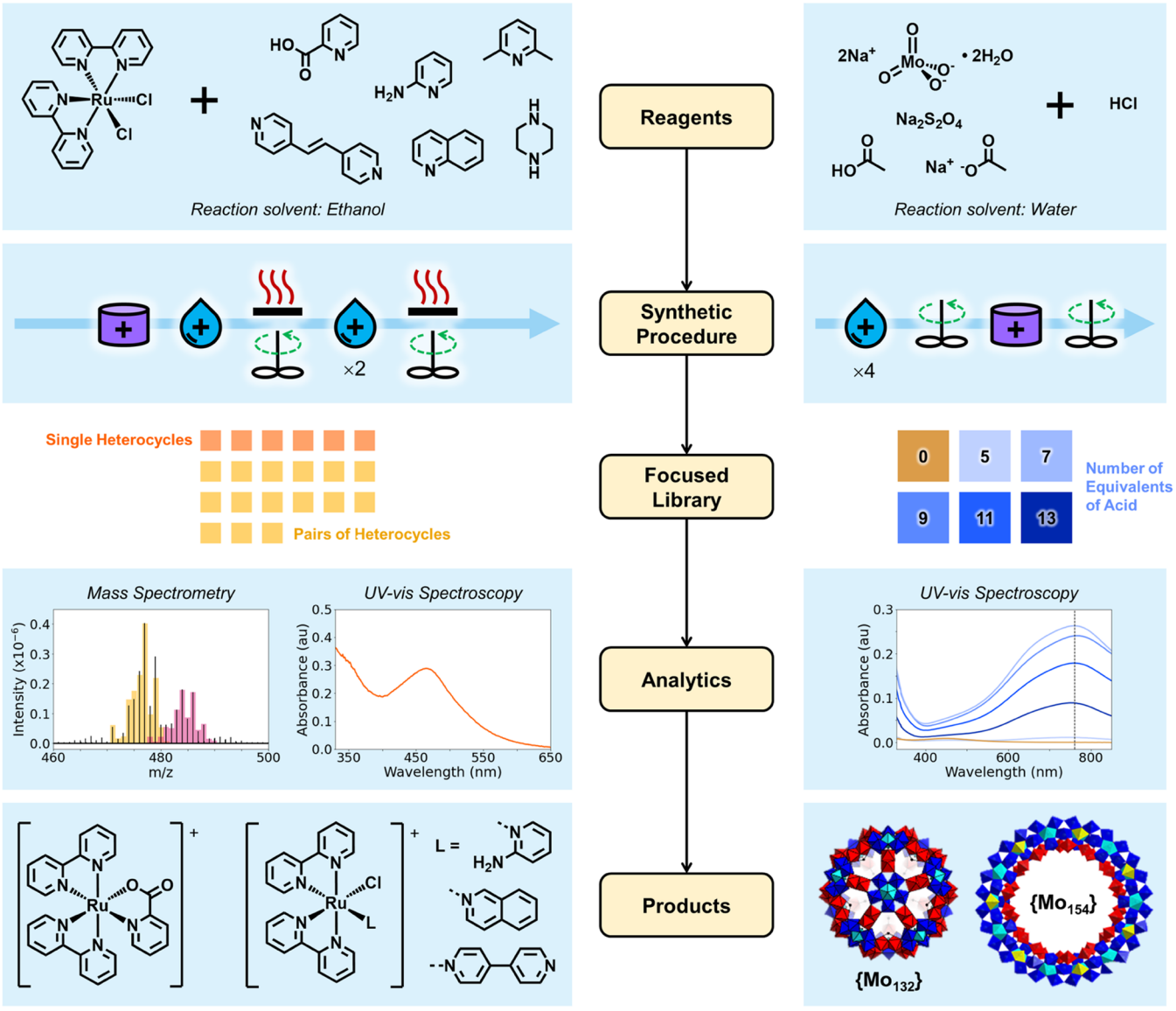


Figure 3. Workflow for the noniterative generation of focused libraries in the Chemputer. (Left) A library in chemical space of $Ru(bipy)_2Cl_2$ derivatives. (Right) A library in process space of reaction compositions leading to the formation of polyoxomolybdates. In each case, the reagents, synthetic procedure, library construction, methods of analysis, and isolated products are shown. Polyoxomolybdate polyhedra are colored according to the structural building blocks represented: $\{Mo\{Mo_5\}\}$, cyan and blue; $\{Mo_1\}$, yellow; $\{Mo_2\}$ in red. Note that $\{Mo_2\}$ building blocks in $\{Mo_{132}\}$ are formed from edge-sharing octahedral, and corner-sharing in $\{Mo_{154}\}$.

isoquinoline, and 2-aminopyridine displaced a single chlorido ligand to form complexes bound through the pyridyl nitrogen. Piperazine showed evidence of new species by MS, with the isotope pattern indicating the presence of ruthenium, but these could not be unambiguously assigned. Pairwise combinations of the *N*-heterocycles were also attempted in a brute-force screen, with MS data implying that only complexes observed in the initial library were formed.

A second experiment varied the number of equivalents of hydrochloric acid in the synthesis of polyoxomolybdates—nanoscale species[40] with interesting self-assembling behavior[41,42] and a variety of applications.[43] When conducted noniteratively, this is effectively a focused library in process space, as opposed to the chemical space exemplified by the previous library (Figure 3, right).

The system chosen required premixing solutions of $Na_2MoO_4{\cdot}2H_2O$, NaOAc, 50% acetic acid in water, and 1 M HCl in one reactor flask, before combining this mixture with solid sodium dithionite as a reducing agent in a second flask. Depending on the pH of the reaction mixture, the system self-assembles to give one of two polyoxometalates, the $\{Mo_{132}\}$ ball or a $\{Mo_{154}\}$ ring. Reactions were conducted using a series of equivalents (0−13) of 1 M $HCl_{(aq)}$. The two species have previously been observed to coexist around pH 2.6,[42] and we also observe this behavior where 5 equivalents of acid are used. The yield of $\{Mo_{132}\}$ was maximized in the absence of acid, and $\{Mo_{154}\}$ with 7 equivalents.

Prospecting Library Generation. Our workflow for prospecting library generation uses an algorithm to direct the exploration of a process space toward areas of the space more likely to produce more novel species. This is achieved through the creation of a metric built on the difference between given experimental spectra and both the starting material and previous experimental spectra. The larger the difference, the larger the

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341

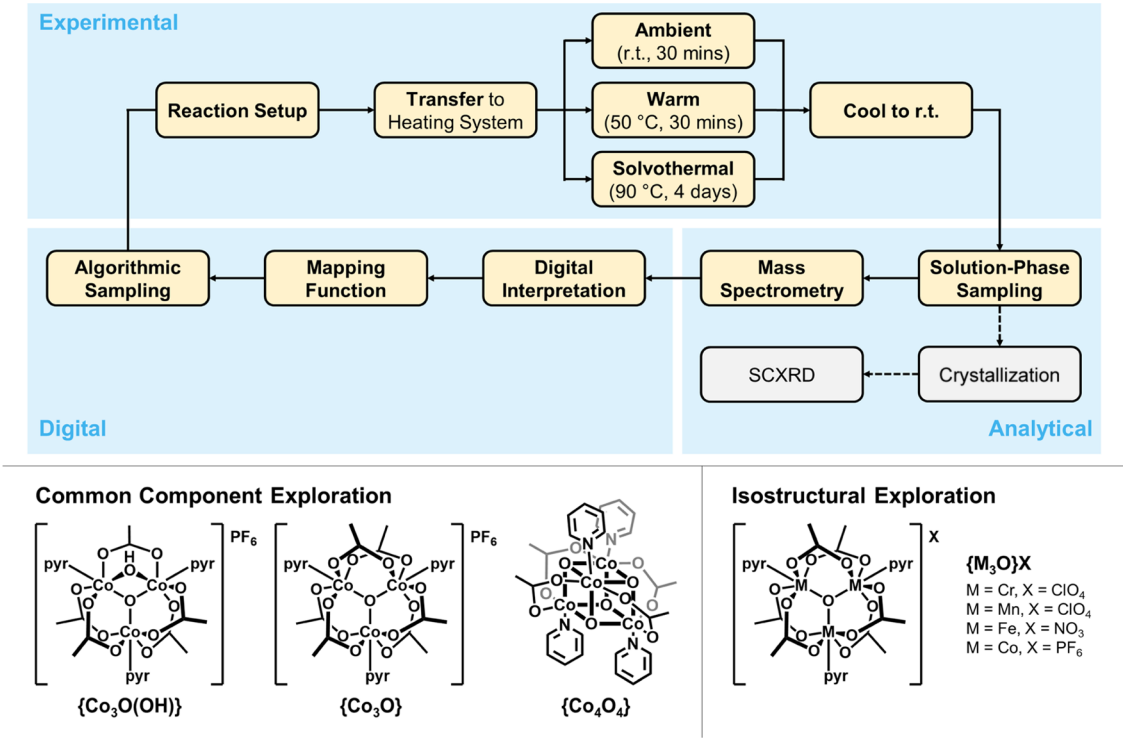


Figure 4. (Top) Workflow for the autonomous exploration of a cluster metal carboxylate search space. Steps are broken down into the Experimental, Analytical, and Digital sections of the workflow. Single-crystal X-ray diffraction (SCXRD) analysis was attempted where crystals of sufficient quality could be isolated but did not inform the sampling of further iterations. (Bottom) Cluster starting materials used in the two explorations to generate the prospecting libraries.

score, and thus the more likely a serendipitous or novel discovery has been made.

Directing the exploration toward the discovery of novelty explicitly should enumerate the chemical space accessible, effectively creating a prospecting library containing these enumerated products. However, unconstrained exploration does not guarantee purity in the systems produced, and so we also propose a method of coalescing and deconvoluting the resulting MS dataset to allow chemical diversity to be screened at the systems level. This removes the requirement to optimize the purification of many systems of varying composition.

We decided to probe cluster metal carboxylates, as such species offer a highly diverse chemical space of theoretical products. The system is highly reconfigurable, with aggregation or decomposition leading to a large space formed by very few building blocks. The combinatorial nature of this chemical space makes reaction outcomes difficult to predict heuristically or by simulation.[2] Cluster metal carboxylates have a wide range of potential applications as biological model compounds and materials.[29,30,32,44−51] The space of possible compounds is expanded through the use of additives with structure-directing effects, which were chosen for their potential to promote higher nuclearity structure formation through templation (lanthanide chloride hydrate salts), bridging between self-contained cluster building blocks (ditopic carboxylic acids), and steric shielding of the metal-oxo core (*N*,*N*′,*N*″-trimethyl-1,4,7-triazacyclononane, TMTACN). Two chemical systems were ultimately used, one involving species formed only from Co(III), acetate, pyridine, and water (the Common Component Exploration), and the other comprising species with the general formula $[M_3O(OAc)_6(pyr)_3]^+$ and a variety of metals, M = Cr, Mn, Fe, Co (the Isostructural Exploration). The exploration proceeded iteratively, following a common pattern from similar experiments, which includes reaction, analysis, and digital interpretation of the analytical data to score each reaction. Finally, suggestion of a subsequent round of experiments is made by a digital agent (Figure 4).[6]

Rather than using the Chemputer in the same form as for the focused libraries, reactions were performed directly in a variation of the sample storage carousel. This increased the throughput of the exploration beyond the four reactor modules to a maximum of 24 simultaneous reactions. The separate Geneva Wheel Platform (GWP) was outfitted to permit direct sample dispensing and stirring,[24] with a batch heater used to enable temperature control of the reaction. Reactions were conducted under three temperatures—ambient conditions for 30 min, 50 °C for 30 min, and solvothermally at 90 °C for 4 days. An iteration comprised 48 reaction compositions each conducted at the three temperatures. Electrospray ionization mass spectrometry (ESI-MS) spectra were recorded for every combination of composition and temperature by sampling the diluted reaction mixtures directly.

MS spectra were then scored according to a bespoke metric designed to quantify the novelty of experiments. The metric comprises two halves: (i) a score of the percentage difference between the five most intense peaks in the spectra of each starting material and the five most intense peaks in each experiment, and (ii) a count of any peaks in the experiment spectrum (above a noise threshold) that do not appear in cluster starting material spectra. Weighting of these peaks by the inverse of their frequency within the dataset means that commonly occurring peaks contribute less to this second measure than do rarer peaks. The scores are recalculated with each iteration, as

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341

the frequency parameter of the second measure must be updated to cover the full dataset fairly.

The two halves of the novelty score were then individually feature-scaled between 0 and 1, to weight them evenly, before averaging to give a final score for each experimental spectrum. These scores were further averaged over the three temperatures. In this way, each of the 48 compositions used in an iteration is given a single numerical score between 0 and 1. Averaging across temperatures runs the risk of obscuring some of the variation in the results due to this variable, but with this methodology validated, steps can be taken to improve on this in future explorations.

The initial iteration of experimental compositions was chosen by Latin Hypercube sampling (LHS) to ensure an even spread of data points over the search space. Subsequent experiments were chosen via Bayesian optimization with a Gaussian process surrogate model as prior (GPBO).[52] The GP model was built from the calculated novelty scores using a combined Matérn and white noise kernel. The optimization used Lower Common Bound (LCB) as the acquisition function to permit easy tuning between exploitation and exploration of the space. Uniform random sampling was also used to generate experiments for one iteration of the Common Component Exploration, to provide a comparison for GPBO. A short primer on each method is available in the SI. The data were evaluated by other offline analytical methods where possible. Any experiments yielding crystalline species were analyzed by SCXRD, and all MS spectra were then subjected to a digital pipeline to deconvolute usable chemical insights from the spectral data (discussed in the next section).

Under ambient and warm (50 °C) conditions, the first and second iterations of the Common Component Exploration afforded a number of crystals that matched species reported to be intermediates in the synthesis of $\{Co_3O\}$ and $\{Co_3O(OH)\}$ starting materials.[53] Two of these were distinct species—$[Co_3O(OH)_2(OAc)_3(pyr)_5](PF_6)_2$ and $[Co_2(OH)_2(OAc)_3(pyr)_4]PF_6$—and a further two methanolysis products. In assessing the significance of these discoveries, it is of note that the optimizer does not bias its data collection to align with any pre-programmed chemical heuristic. Although these complexes are known in the literature, they are completely new to this system and are scored accordingly.

Under solvothermal conditions, the first iteration of the Common Component Exploration afforded crystals corresponding to a family of three acetic anhydride complexes of Co(III) (Figure 5). All of these appear to be new to the literature, with only a single example of an acetic anhydride complex (of Ti(IV))[54] found in the Cambridge Structural Database. Two of the complexes formed appear to be of structural types proposed from *in situ* analysis of anhydride complexes of other metal ions,[55−58] and the third, $[Co(AcOAc)_2(OAc)]^{2+}$, has not been proposed previously with these ligands.[59,60] Counteranions could not be unambiguously assigned for this species; however, ^{1}H NMR data show agreement with the crystal structure of the cationic portion. Acetic anhydride is believed to form via condensation of the acetic acid and NMR analysis implies that this is promoted in the presence of cobalt in either the +II or +III oxidation states. Cyano ligands are hypothesized to originate from the acetonitrile solvent, with metal-mediated C−C cleavage of acetonitrile previously reported, including an example with a cationic Rh(III) complex.[61−67]

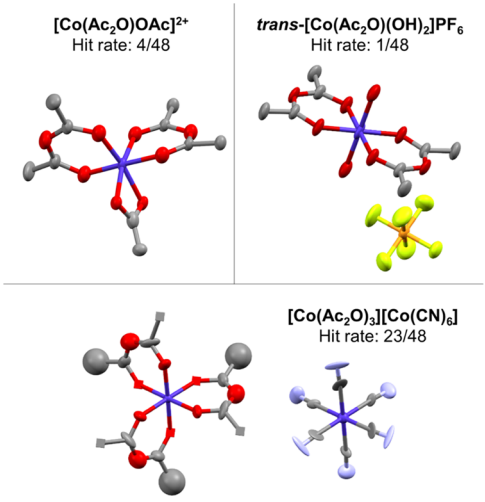


Figure 5. Cobalt(III) anhydride complexes isolated from solvothermal (90 °C for 4 days) samples in the first iteration of the Common Component Exploration. "Hit rate" refers to the number of experiments affording crystals of the material from the first iteration. Ellipsoid representations of Co, dark blue; C, gray; O, red; N, light blue; P, orange; F, light green.

The Isostructural Exploration afforded no crystals under ambient or warm conditions except for methanolysis products. Under solvothermal conditions, however, almost all samples afforded crystals of $[M(OMe)_2(OAc)]_{10}$, $\{M_{10}\}$. Under comparable conditions with only a single metal cluster precursor, only the Fe(III) system gave an analogous species,[68] although $\{Cr_{10}\}$ rings are also reported in the literature.[69] An IR spectrum of the crystals did not match that of the $\{Fe_{10}\}$ ring precisely, with notable differences in the M−O stretching region, implying a multimetallic composition.

Deconvolution of MS Data from the Prospecting Library. Deconvolution was achieved by taking the list of unique peaks for each experiment at a given temperature, converting these to an ordered list indicating the presence or absence of a peak at a given m/z with a binary notation, and the array of these "barcodes" subjected to non-negative matrix factorization (NMF), to afford a lower-rank approximation of the original barcode array (Figure 6). The rank is determined by screening ranks from 2 to the number of experiments in the barcode array and using as a threshold the rank at which reconstruction error becomes minimized. In this way, the coefficient matrix corresponds to groupings of m/z values that commonly occur together—effectively, fingerprints corresponding to specific product distributions. The feature matrix provides the composition of the MS spectrum for each experiment in terms of these proposed product distributions. We refer to the proposed product distributions as "Archetypes."

NMF was chosen over similar methods (e.g., principal component analysis, PCA) due to the ability to reconstruct the approximated original dataset through only additive combination, which the original paper from Lee and Seung[70] highlights as being "compatible with the intuitive notion of combining parts to form a whole". This leads to more intuitively interpretable approximated matrices. While NMF has been applied to the interpretation of MS spectra before (including the deconvolution of overlapping spectra,[71] the interpretation of large datasets,[72] and unsupervised classification in mass spectrometry imaging[73,74]), we are not aware of any literature

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341

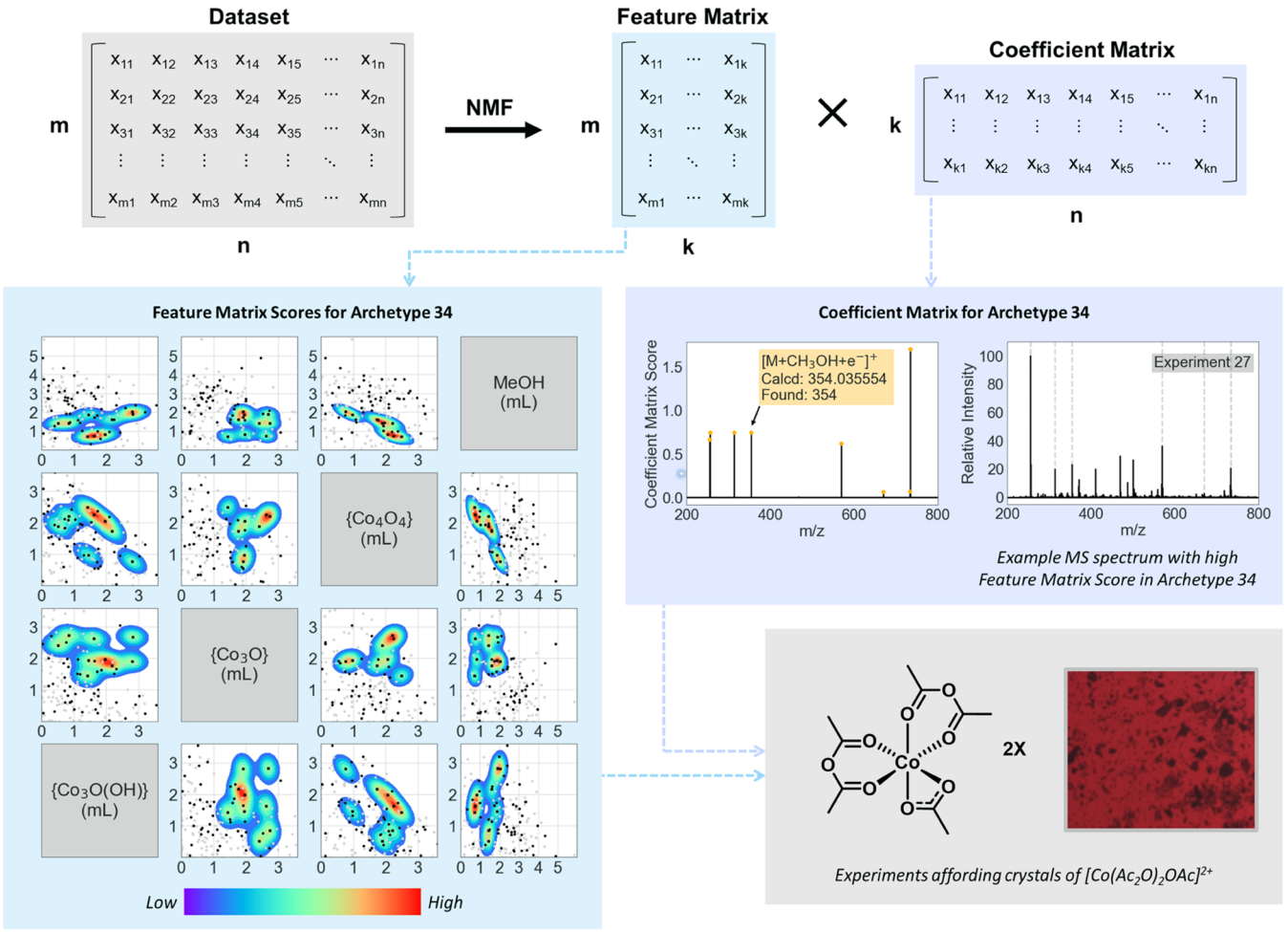


Figure 6. (Top) Overview of the NMF process, where *m* is the experiment number, *n* corresponds to each unique peak observed throughout the dataset, and *k* is effectively groupings of commonly occurring peaks in the *n* set—i.e., MS archetypes. (Bottom left) Truncated matrix plot of the search space. Data points indicate conditions of each experiment. A kernel density estimation, weighted by the feature matrix score for the $[Co(Ac_2O)OAc]^{2+}$-related archetype, indicates where this product distribution is located within the process space. (Mid right) Corresponding coefficient matrix scores and comparison with experimental MS data.

leveraging the beneficial properties of the technique in a discovery context.

Examination of the feature matrix scores for those experiments in the Common Component Exploration yielding crystals of $[Co(AcOAc)_2(OAc)]^{2+}$ can be used to demonstrate that the archetypes have chemical significance. All four of these experiments show high scores in a particular archetype (with some feature matrix variation unique to each experiment). Examination of the corresponding entry in the coefficient matrix indicates that this archetype corresponds to peaks that may be assigned to the crystallized species, notably $m/z = 354$, assigned as $[M + CH_3OH + e^-]^+$. The presence of this archetype in samples from later iterations indicates the likely presence of $[Co(AcOAc)_2(OAc)]^{2+}$ in these experiments, despite no SCXRD-quality crystals being collected.

The feature matrix scores for a given archetype can be used to draw conclusions about the areas of the process space (i.e., the synthetic conditions) that afford a given product distribution. This is represented for the discriminatory $[Co(AcOAc)_2(OAc)]^{2+}$ archetype in Figure 6 as a kernel density estimation weighted by the archetype's feature matrix score for each experiment. The archetype with the largest percentage contribution to the summed feature matrix scores of a given experiment may be readily calculated, giving potential insights into the purity of product mixtures—information that has value for further synthesis but is otherwise difficult to obtain from MS data.

The NMF approximation may also give insight into the efficacy of our exploration approach. Comparing the increase in archetypes discovered between the initial Common Component Exploration Run 1 and the GPBO-directed Run 2 at all temperatures, against the increase between Run 1 and the randomly-sampled Run 3, we observe that GPBO discovers more archetypes (and thus more product distributions) than random sampling. The percentage of archetypes discovered in Run 2 vs Run 3 that are shared discoveries varies between 50 and 80% depending on temperature, indicating that, to some extent, different areas of the search space are targeted.

If NMF approximations are created not from datasets of all experiments in an exploration at a given temperature, but from experiments from each iteration at each temperature, we observe that GPBO-directed Runs 2 and 4 tend to require a higher rank to minimize reconstruction error. This implies a greater chemical diversity discovered during these GPBO runs. In the Isostructural Exploration, this trend is not readily apparent, which we suggest is due to an overall lower chemical diversity

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341

present in the search space. Thus, GPBO appears to be a better method of exploration than random sampling for the discovery of novelty in process space, in terms of the number and diversity of new product distributions discovered.

CONCLUSIONS

In this work, we have demonstrated the production of focused and prospecting libraries[8] in coordination chemistry using programmable chemical reaction platforms like the Chemputer.[16] Such platforms can generalize our methods beyond the examples shown due to the use of a programming approach based on the abstraction of chemical syntheses.

Generalizability is demonstrated through the initial TOS of three coordination complexes and the metastable polymorphs of calcium carbonate. Key facets of inorganic synthesis are represented, including purification by recrystallization, self-assembly of multinuclear species, stereoselective synthesis, and preparation of thermodynamically unstable phases. Parallelization was then used to generate two focused libraries. The first allowed the synthetic enumeration of a number of $Ru(bipy)_2Cl_2$ derivatives, and the second demonstrated the relationship between synthesis conditions and yield of two polyoxomolybdates. In this way, we show that focused libraries can be developed in both chemical and process space by our system.

Further to this, we provide a workflow for the forward synthesis of a prospecting library of coordination entities that links the product distributions discovered to the areas of process space they occur within. This allows us to screen for the diversity present in the search space while overcoming the common criticism of DOS as not leading to the development of synthetic chemistry.

Our algorithmic explorer has no knowledge of chemistry and is driven toward the areas of the search space showing the rarest results by a novelty quantification metric calculated from MS data. We demonstrate that this method is superior to random sampling—building on previous reports of accelerated serendipity through random choice of reactions[5] to "engineered" serendipity through algorithmic direction. We propose that the resultant dataset of non-starting material peaks can be used to isolate product distributions and demonstrate the validity of this for a species characterized orthogonally by SCXRD. We are able to discover and direct our exploration from the systems level, without losing time, effort, and resource to the purification of many uninteresting results.

Our workflow for screening diversity is generalizable to any system, that can be analyzed reliably by MS, rather than optimized toward a particular application as for DOS. While we chose to direct our exploration by quantifying the novelty of the result, the mapping function could be changed to assay for a particular function or behavior. Central to the diversity discovered is the choice of process space, and further work will aim to highlight considerations and heuristics to allow chemists to best formulate these exploration problems. Future work should also focus on increasing the complexity of the synthetic processes that can be screened (effectively expanding the dimensionality of the process space) to increase the complexity of the product species accessible.

ASSOCIATED CONTENT

Data Availability Statement

Code and related data files are available from GitHub at: github.com/croningp/DigitalCoordinationChemistry.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c11066.

Further discussion of platform construction, data science workflows, and experimental data (PDF)

Accession Codes

CCDC 2211731−2211733 contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 1223 336033.

AUTHOR INFORMATION

Corresponding Author

Leroy Cronin − *School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom;* orcid.org/0000-0001-8035-5757; Email: Leroy.Cronin@glasgow.ac.uk

Authors

Daniel J. Kowalski − *School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom*

Catriona M. MacGregor − *School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom*

De-Liang Long − *School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom;* orcid.org/0000-0003-3241-2379

Nicola L. Bell − *School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom;* orcid.org/0000-0002-7497-9667

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c11066

Author Contributions

The manuscript was written and experimental work undertaken through contribution from all authors.

Funding

The authors thank the following funders: EPSRC (grant nos. EP/H024107/1, EP/J00135X/1, EP/J015156/1, EP/K021966/1, EP/K023004/1, EP/L023652/1), ERC (project 670467 SMART-POM).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Jennifer Mathieson and Maria Diana Castro Spencer for their help with MS measurement and interpretation, Patricia Keating at the University of Strathclyde for MALDI measurements, and Daniel Salley and Graham Keenan for the construction of and instruction in the GWP. For their insights, thanks also go to Jim McIver, Robert Pow, and Christoph Busche (synthesis), as well as S. Hessam M. Mehr, Alexander Hammer, and Yibin Jiang (data science).

ABBREVIATIONS

DOS	diversity-oriented synthesis
ESI	electrospray ionization
GWP	Geneva Wheel Platform
GPBO	Gaussian process Bayesian optimization
LCB	lower common bound
MLCT	metal ligand charge transfer

LHS	Latin hypercube sampling
MS	mass spectrometry
NMF	non-negative matrix factorization
NMR	nuclear magnetic resonance
SAR	structure−activity relationship
SCXRD	single-crystal X-ray diffraction
TMTACN	1,4,7-trimethyl-1,4,7-triazacyclononane
TOS	target-oriented synthesis
UV−vis	ultraviolet−visible (spectroscopy)

REFERENCES

(1) Schreiber, S. L. Target-Oriented and Diversity-Oriented Organic Synthesis in Drug Discovery. *Science* **2000**, *287*, 1964−1969.
(2) Winpenny, R. E. P. Serendipitous assembly of polynuclear cage compounds. *J. Chem. Soc., Dalton Trans.* **2002**, 1−10.
(3) Rulev, A. Y. Serendipity or the art of making discoveries. *New J. Chem.* **2017**, *41*, 4262−4268.
(4) García, P. Discovery by serendipity: a new context for an old riddle. *Found. Chem.* **2009**, *11*, 33−42.
(5) McNally, A.; Prier, C. K.; MacMillan, D. W. C. Discovery of an α-Amino C−H Arylation Reaction Using the Strategy of Accelerated Serendipity. *Science* **2011**, *334*, 1114−1117.
(6) Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous discovery in the chemical sciences part I: Progress. *Angew. Chem., Int. Ed.* **2020**, *59*, 22858−22893.
(7) Shekar, V.; Yu, V.; Garcia, B. J.; Gordon, D. B.; Moran, G. E.; Blei, D. M.; Roch, L. M.; García-Durán, A.; Najeeb, M. A.; Zeile, M.; Nega, P. W.; Li, Z.; Kim, M. A.; Chan, E. M.; Norquist, A. J.; Friedler, S.; Schrier, J. Serendipity based recommender system for perovskites material discovery: balancing exploration and exploitation across multiple models *ChemRxiv* 2022. version 2. DOI: 10.26434/chemrxiv-2022-l1wpf.
(8) Spaller, M. R.; Burger, M. T.; Fardis, M.; Bartlett, P. A. Synthetic strategies in combinatorial chemistry. *Curr. Opin. Chem. Biol.* **1997**, *1*, 47−53.
(9) Burke, M. D.; Schreiber, S. L. A Planning Strategy for Diversity-Oriented Synthesis. *Angew. Chem., Int. Ed.* **2004**, *43*, 46−58.
(10) Galloway, W. R. J. D.; Isidro-Llobet, A.; Spring, D. R. Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. *Nat. Commun.* **2010**, *1*, No. 80.
(11) Nielsen, T. E.; Schreiber, S. L. Towards the Optimal Screening Collection: A Synthesis Strategy. *Angew. Chem., Int. Ed.* **2008**, *47*, 48−56.
(12) Comer, E.; Rohan, E.; Deng, L.; Porco, J. A. An Approach to Skeletal Diversity Using Functional Group Pairing of Multifunctional Scaffolds. *Org. Lett.* **2007**, *9*, 2123−2126.
(13) Schreiber, S. L. Molecular diversity by design. *Nature* **2009**, *457*, 153−154.
(14) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23*, 1213−1242.
(15) Lenci, E.; Trabocchi, A. Diversity-Oriented Synthesis and Chemoinformatics: A Fruitful Synergy towards Better Chemical Libraries. *Eur. J. Org. Chem.* **2022**, *2022*, No. e202200575.
(16) Gromski, P. S.; Granda, J. M.; Cronin, L. Universal Chemical Synthesis and Discovery with 'The Chemputer'. *Trends Chem.* **2020**, *2*, 4−12.
(17) Nandy, A.; Duan, C.; Taylor, M. G.; Liu, F.; Steeves, A. H.; Kulik, H. J. Computational Discovery of Transition-metal Complexes: From High-throughput Screening to Machine Learning. *Chem. Rev.* **2021**, *121*, 9927−10000.
(18) Fernández-Figueiras, A.; Lucio-Martínez, F.; Munín-Cruz, P.; Ortigueira, J. M.; Polo-Ces, P.; Reigosa, F.; Pereira, M. T.; Vila, J. M. From Chemical Serendipity to Translational Chemistry: New Findings in the Reactivity of Palladacycles. *ChemistryOpen* **2018**, *7*, 754−763.
(19) Guidi, M.; Seeberger, P. H.; Gilmore, K. How to approach flow chemistry. *Chem. Soc. Rev.* **2020**, *49*, 8910−8932.
(20) Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J. M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P. J.; Angelone, D.; Cronin, L. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **2019**, *363*, No. eaav2211.
(21) Angelone, D.; Hammer, A. J. S.; Rohrbach, S.; Krambeck, S.; Granda, J. M.; Wolf, J.; Zalesskiy, S.; Chisholm, G.; Cronin, L. Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. *Nat. Chem.* **2021**, *13*, 63−69.
(22) Bornemann-Pfeiffer, M.; Wolf, J.; Meyer, K.; Kern, S.; Angelone, D.; Leonov, A.; Cronin, L.; Emmerling, F. Standardization and Control of Grignard Reactions in a Universal Chemical Synthesis Machine using online NMR. *Angew. Chem., Int. Ed.* **2021**, *60*, 23202−23206.
(23) Rohrbach, S.; Šiaučiulis, M.; Chisholm, G.; Pirvan, P.-A.; Saleeb, M.; Mehr, S. H. M.; Trushina, E.; Leonov, A. I.; Keenan, G.; Khan, A.; Hammer, A.; Cronin, L. Digitization and validation of a chemical synthesis literature database in the ChemPU. *Science* **2022**, *377*, 172−180.
(24) Salley, D. S.; Keenan, G. A.; Long, D.-L.; Bell, N. L.; Cronin, L. A Modular Programmable Inorganic Cluster Discovery Robot for the Discovery and Synthesis of Polyoxometalates. *ACS Cent. Sci.* **2020**, *6*, 1587−1593.
(25) Mehr, S. H. M.; Craven, M.; Leonov, A. I.; Keenan, G.; Cronin, L. A universal system for digitization and automatic execution of the chemical synthesis literature. *Science* **2020**, *370*, 101−108.
(26) Rankin, D. W. H.; Mitzel, N. W.; Morrison, C. A. *Structural Methods in Molecular Inorganic Chemistry*; John Wiley & Sons, Ltd., 2013; pp 289−298.
(27) Arias-Rotondo, D. M.; McCusker, J. K. The photophysics of photoredox catalysis: a roadmap for catalyst design. *Chem. Soc. Rev.* **2016**, *45*, 5803−5820.
(28) Shaw, M. H.; Twilton, J.; MacMillan, D. W. C. Photoredox Catalysis in Organic Chemistry. The. *J. Org. Chem.* **2016**, *81*, 6898−6926.
(29) Christou, G. Manganese carboxylate chemistry and its biological relevance. *Acc. Chem. Res.* **1989**, *22*, 328−335.
(30) Aromí, G.; Brechin, E. K. Synthesis of 3d Metallic Single-Molecule Magnets. In *Single-Molecule Magnets and Related Phenomena*; Winpenny, R., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; pp 1−67.
(31) Viciano-Chumillas, M.; Tanase, S.; Mutikainen, I.; Turpeinen, U.; de Jongh, L. J.; Reedijk, J. Mononuclear Manganese(III) Complexes as Building Blocks for the Design of Trinuclear Manganese Clusters: Study of the Ligand Influence on the Magnetic Properties of the $[Mn_3(\mu^3\text{-}O)]^{7+}$ Core. *Inorg. Chem.* **2008**, *47*, 5919−5929.
(32) Bush, J. B.; Finkbeiner, H. Oxidation reactions of manganese(III) acetate. II. Formation of γ-lactones from olefins and acetic acid. *J. Am. Chem. Soc.* **1968**, *90*, 5903−5905.
(33) Vincent, J. B.; Chang, H. R.; Folting, K.; Huffman, J. C.; Christou, G.; Hendrickson, D. N. Preparation and physical properties of trinuclear oxo-centered manganese complexes of general formulation $[Mn_3O(O_2CR)_6L_3]^{0,+}$ (R = methyl or phenyl; L = a neutral donor group) and the crystal structures of $[Mn_3O(O_2CMe)_6(pyr)_3](pyr)$ and $[Mn_3O(O_2CPh)_6(pyr)_2(H_2O)]\cdot 0.5MeCN$. *J. Am. Chem. Soc.* **1987**, *109*, 5703−5711.
(34) WHO. *World Health Organisation Model List of Essential Medicines - 22nd List*; World Health Organisation: Geneva, 2021.
(35) Alderden, R. A.; Hall, M. D.; Hambley, T. W. The Discovery and Development of Cisplatin. *J. Chem. Educ.* **2006**, *83*, 728.
(36) Dhara, S. C. A rapid method for the synthesis of *cis*-$[PtCl_2(NH_3)_2]$ *Indian J. Chem.* 1970, *193*.
(37) Boreham, C. J.; Broomhead, J. A.; Fairlie, D. P. A ^{195}Pt and ^{15}N N.M.R. study of the anticancer drug, cis-diammine-dichloroplatinum-

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341

(II), and its hydrolysis and oligomerization products. *Aust. J. Chem.* **1981**, *34*, 659−664.
(38) Sarkar, A.; Mahapatra, S. Synthesis of All Crystalline Phases of Anhydrous Calcium Carbonate. *Cryst. Growth Des.* **2010**, *10*, 2129−2135.
(39) Liu, R.; Li, X.; Lam, K. S. Combinatorial chemistry in drug discovery. *Curr. Opin. Chem. Biol.* **2017**, *38*, 117−126.
(40) Miras, H. N.; Yan, J.; Long, D.-L.; Cronin, L. Engineering polyoxometalates with emergent properties. *Chem. Soc. Rev.* **2012**, *41*, 7403−7430.
(41) Miras, H. N.; Cooper, G. J. T.; Long, D.-L.; Bögge, H.; Müller, A.; Streb, C.; Cronin, L. Unveiling the Transient Template in the Self-Assembly of a Molecular Oxide Nanowheel. *Science* **2010**, *327*, 72.
(42) Miras, H. N.; Mathis, C.; Xuan, W.; Long, D.-L.; Pow, R.; Cronin, L. Spontaneous formation of autocatalytic sets with self-replicating inorganic metal oxide clusters. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 10699.
(43) Katsoulis, D. E. A Survey of Applications of Polyoxometalates. *Chem. Rev.* **1998**, *98*, 359−388.
(44) Sessoli, R.; Tsai, H. L.; Schake, A. R.; Wang, S.; Vincent, J. B.; Folting, K.; Gatteschi, D.; Christou, G.; Hendrickson, D. N. High-spin molecules: $[Mn_{12}O_{12}(O_2CR)_{16}(H_2O)_4]$. *J. Am. Chem. Soc.* **1993**, *115*, 1804−1816.
(45) Reisner, E.; Telser, J.; Lippard, S. J. A Planar Carboxylate-Rich Tetrairon(II) Complex and Its Conversion to Linear Triiron(II) and Paddlewheel Diiron(II) Complexes. *Inorg. Chem.* **2007**, *46*, 10754−10770.
(46) Taft, K. L.; Delfs, C. D.; Papaefthymiou, G. C.; Foner, S.; Gatteschi, D.; Lippard, S. J. $[Fe(OMe)_2(O_2CCH_2Cl)]_{10}$, a Molecular Ferric Wheel. *J. Am. Chem. Soc.* **1994**, *116*, 823−832.
(47) Heiba, E. I.; Dessau, R.; Rodewald, P. Oxidation by metal salts. X. One-step synthesis of γ-lactones from olefins. *J. Am. Chem. Soc.* **1974**, *96*, 7977−7981.
(48) Dunlap, N. K.; Sabol, M. R.; Watt, D. S. Oxidation of enones to α′-acetoxyenones using manganese triacetate. *Tetrahedron Lett.* **1984**, *25*, 5839−5842.
(49) Melikyan, G. G. Carbon−Carbon Bond-Forming Reactions Promoted by Trivalent Manganese. *Org. React.* **1996**, 427−675.
(50) Ghosh, T.; Maayan, G. Efficient Homogeneous Electrocatalytic Water Oxidation by a Manganese Cluster with an Overpotential of Only 74 mV. *Angew. Chem., Int. Ed.* **2019**, *58*, 2785−2790.
(51) Chen, R.; Yan, Z.-H.; Kong, X.-J. Recent Advances in First-Row Transition Metal Clusters for Photocatalytic Water Splitting. *ChemPhotoChem* **2020**, *4*, 157−167.
(52) Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; Freitas, Nd. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148−175.
(53) Sumner, C. E. Interconversion of dinuclear and oxo-centered trinuclear cobaltic acetates. *Inorg. Chem.* **1988**, *27*, 1320−1327.
(54) Viard, B.; Poulain, M.; Grandjean, D.; Amaudrut, J. PREPARATION AND STRUCTURE DETERMINATION OF COMPLEXES OF ACETIC-ANHYDRIDE - REACTION OF ACETIC-ANHYDRIDE WITH CHLORIDES OF ELEMENTS IN GROUP-IV AND GROUP-V. *J. Chem. Res., Synop.* **1983**, 84−85.
(55) Gutmann, V.; Schmidt, H. Das Perchloration als koordinierendes Anion in kristallisierten Co(II)- und Ni(II)-Solvaten. *Monatsh. Chem.* **1974**, *105*, 653−665.
(56) Murase, M.; Kotani, E.; Okazaki, K.; Tobinaga, S. Application of Iron(III) Complexes, Tris(2, 2′-bipyridyl)iron(III) Perchlorate and Some Iron(III) Solvates, for Oxidative Aryl-Aryl Coupling Reactions. *Chem. Pharm. Bull.* **1986**, *34*, 3159−3165.
(57) Kotani, E.; Kobayashi, S.; Adachi, M.; Tsujioka, T.; Nakamura, K.; Tobinaga, S. Synthesis of Oxazoles by the Reaction of Ketones with Iron(III) Solvates of Nitriles. *Chem. Pharm. Bull.* **1989**, *37*, 606−609.
(58) Il'in, E. G.; Kovalev, V. V. Coordination modes and reaction of acetic anhydride with tantalum pentafluoride. *Russ. J. Inorg. Chem.* **2014**, *59*, 111−114.
(59) Qu, Y.; Zhu, H.-L.; You, Z.-L.; Tan, M.-Y. Synthesis and Characterization of the Complexes of Pentane-2,4-dione with Nickel(II) and Cobalt(III): $[Ni(acac)_2]\cdot 0.5CH_3OH$ and $[Co(acac)_2NO_3]\cdot 2H_2O$ (acac = Pentane-2,4-dione). *Molecules* **2004**, *9*, 949−956.
(60) Cakić, S. M.; Nikolić, G. S.; Stamenković, J. V.; Konstantinović, S. S. Physico-chemical characterization of mixed-ligand complexes of Mn(III) based on the acetylacetonate and maleic acid and its hydroxylamine derivative. *Acta Period. Technol.* **2005**, 91−98.
(61) Luo, F.-H.; Chu, C.-I.; Cheng, C.-H. Nitrile-Group Transfer from Solvents to Aryl Halides. Novel Carbon−Carbon Bond Formation and Cleavage Mediated by Palladium and Zinc Species. *Organometallics* **1998**, *17*, 1025−1030.
(62) Marlin, D. S.; Olmstead, M. M.; Mascharak, P. K. Heterolytic Cleavage of the C−C Bond of Acetonitrile with Simple Monomeric Cu(II) Complexes: Melding Old Copper Chemistry with New Reactivity. *Angew. Chem., Int. Ed.* **2001**, *40*, 4752−4754.
(63) Taw, F. L.; White, P. S.; Bergman, R. G.; Brookhart, M. Carbon−Carbon Bond Activation of R−CN (R = Me, Ar, iPr, tBu) Using a Cationic Rh(III) Complex. *J. Am. Chem. Soc.* **2002**, *124*, 4192−4193.
(64) Nakazawa, H.; Kawasaki, T.; Miyoshi, K.; Suresh, C. H.; Koga, N. C−C Bond Cleavage of Acetonitrile by a Carbonyl Iron Complex with a Silyl Ligand. *Organometallics* **2004**, *23*, 117−126.
(65) Ateşin, T. A.; Li, T.; Lachaize, S.; Brennessel, W. W.; García, J. J.; Jones, W. D. Experimental and Theoretical Examination of C−CN and C−H Bond Activations of Acetonitrile Using Zerovalent Nickel. *J. Am. Chem. Soc.* **2007**, *129*, 7562−7569.
(66) Song, R.-J.; Wu, J.-C.; Liu, Y.; Deng, G.-B.; Wu, C.-Y.; Wei, W.-T.; Li, J.-H. Copper-Catalyzed Oxidative Cyanation of Aryl Halides with Nitriles Involving Carbon−Carbon Cleavage. *Synlett* **2012**, *23*, 2491−2496.
(67) Pan, C.; Jin, H.; Xu, P.; Liu, X.; Cheng, Y.; Zhu, C. Copper-Mediated Direct C2-Cyanation of Indoles Using Acetonitrile as the Cyanide Source. *J. Org. Chem.* **2013**, *78*, 9494−9498.
(68) Parsons, S.; Solan, G. A.; Winpenny, R. E. P.; Benelli, C. Ferric Wheels and Cages: Decanuclear Iron Complexes with Carboxylato and Pyridonato Ligands. *Angew. Chem., Int. Ed.* **1996**, *35*, 1825−1828.
(69) McInnes, E. J. L.; Anson, C.; Powell, A. K.; Thomson, A. J.; Poussereau, S.; Sessoli, R. Solvothermal synthesis of $[Cr_{10}(\mu\text{-}O_2CMe)_{10}(\mu\text{-}OR)_{20}]$ 'chromic wheels' with antiferromagnetic (R = Et) and ferromagnetic (R = Me) Cr(iii)···Cr(iii) interactions. *Chem. Commun.* **2001**, 89−90.
(70) Lee, D. D.; Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788−791.
(71) Gao, H.-T.; Li, T.-H.; Chen, K.; Li, W.-G.; Bi, X. Overlapping spectra resolution using non-negative matrix factorization. *Talanta* **2005**, *66*, 65−73.
(72) Trindade, G. F.; Abel, M.-L.; Watts, J. F. Non-negative matrix factorisation of large mass spectrometry datasets. *Chemom. Intell. Lab. Syst.* **2017**, *163*, 76−85.
(73) Nijs, M.; Smets, T.; Waelkens, E.; De Moor, B. A mathematical comparison of non-negative matrix factorization related methods with practical implications for the analysis of mass spectrometry imaging data. *Rapid Commun. Mass Spectrom.* **2021**, *35*, No. e9181.
(74) Xiong, X.-C.; Fang, X.; Ouyang, Z.; Jiang, Y.; Huang, Z.-J.; Zhang, Y.-K. Feature Extraction Approach for Mass Spectrometry Imaging Data Using Non-negative Matrix Factorization. *Chin. J. Anal. Chem.* **2012**, *40*, 663−669.

https://doi.org/10.1021/jacs.2c11066
J. Am. Chem. Soc. 2023, 145, 2332−2341