ACS central science

http://pubs.acs.org/journal/acscii

Research Article

Open-Source Chromatographic Data Analysis for Reaction Optimization and Screening

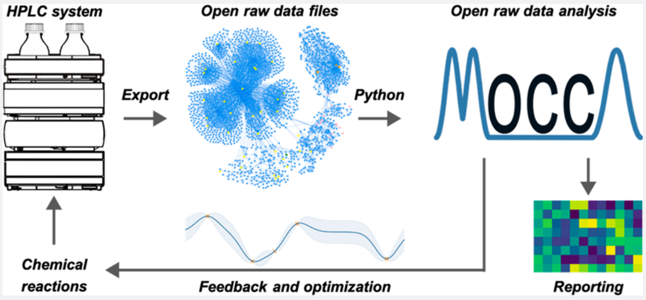
Christian P. Haas, Maximilian Lübbesmeyer, Edward H. Jin, Matthew A. McDonald, Brent A. Koscher, Nicolas Guimond, Laura Di Rocco, Henning Kayser, Samuel Leweke, Sebastian Niedenführ, Rachel Nicholls, Emily Greeves, David M. Barber, Julius Hillenbrand,* Giulio Volpin,* and Klavs F. Jensen*

Cite This: *ACS Cent. Sci.* 2023, 9, 307−317

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Automation and digitalization solutions in the field of small molecule synthesis face new challenges for chemical reaction analysis, especially in the field of high-performance liquid chromatography (HPLC). Chromatographic data remains locked in vendors' hardware and software components, limiting their potential in automated workflows and data science applications. In this work, we present an open-source Python project called MOCCA for the analysis of HPLC−DAD (photodiode array detector) raw data. MOCCA provides a comprehensive set of data analysis features, including an automated peak deconvolution routine of known signals, even if overlapped with signals of unexpected impurities or side products. We highlight the broad applicability of MOCCA in four studies: (i) a simulation study to validate MOCCA's data analysis features; (ii) a reaction kinetics study on a Knoevenagel condensation reaction demonstrating MOCCA's peak deconvolution feature; (iii) a closed-loop optimization study for the alkylation of 2-pyridone without human control during data analysis; (iv) a well plate screening of categorical reaction parameters for a novel palladium-catalyzed cyanation of aryl halides employing *O*-protected cyanohydrins. By publishing MOCCA as a Python package with this work, we envision an open-source community project for chromatographic data analysis with the potential of further advancing its scope and capabilities.

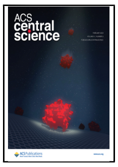


1. INTRODUCTION

Synthetic chemistry enables discovery of new chemical reactivity, access to new molecules of interest, and development of corresponding chemical processes. Ever more demanding regulatory and sustainability requirements on small molecules' synthesis and development make this endeavor complex and cost-intensive.[1−5] Increasing emphasis is given to automation and digitalization in synthetic chemistry to address today's complex challenges while decreasing development time and cost.[6−8] Automation approaches aim to facilitate chemical synthesis while increasing its safety, robustness, and efficiency.[9−11] Digitalization approaches focus mainly on reducing the number of synthetic experiments until a given goal is reached by predicting experimental outcomes or molecular properties. For that, data science techniques are applied to existing data for experimental design and decision making.[12−15] In both areas, generalization to the complexity and diversity of chemical reaction processes remains the main challenge. As stated by Hein and co-workers,[16] "automation isn't automatic," and automated experimental setups are too often tailored to a given synthetic problem.[17−19] Digitalized approaches toward machine learning design algorithms suffer from a highly unstructured data foundation in literature, since data was not collected and reported with data science applications in mind.[20] Therefore, recent approaches focus on the community-based standardization of synthetic lab data (Open Reaction Database), increasing the robustness of experimental protocols against noise, or the augmentation of literature data by systematic experiments performed by automated machines.[21−24]

Interestingly, chemical reaction analysis has received less attention in recent automation and digitalization efforts despite its importance for the overall synthetic process.[25] Analytical raw data generation and analysis remain locked in vendor-specific proprietary hardware and software components, especially in the field of high-performance liquid chromatography (HPLC), a standard analytical method for chemical reaction analysis. Most HPLC systems in academic and

Received: September 5, 2022
Published: February 9, 2023



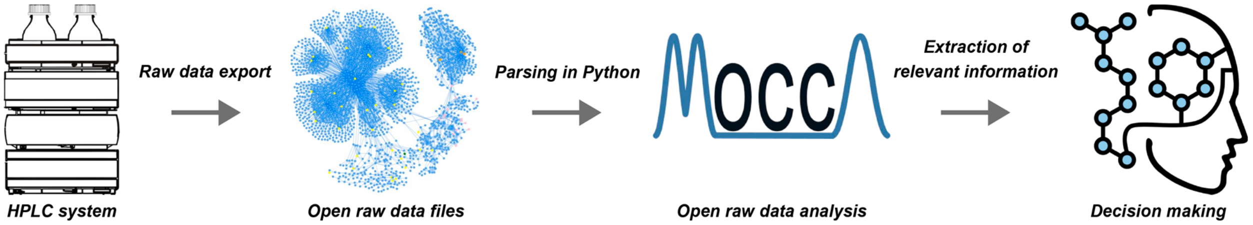


Figure 1. Proposed analytical workflow starting with HPLC systems controlled by vendor-specific software. HPLC−DAD raw data are exported in nonproprietary and open data formats, preferably, a metadata-enriched standardized format (Allotrope) implementing FAIR data principles. After parsing in Python, HPLC−DAD data sets are analyzed in context to each other by MOCCA. From the analysis results, structured data sets are generated for data-based decision making.

industrial research laboratories are equipped with photodiode array detectors (DAD) that record full UV−Vis spectra at every chromatogram time point. For analysis, the dimensionality of the HPLC−DAD data is classically reduced to chromatograms by vendor data analysis software, i.e., absorbance at a single wavelength as a function of retention time. In such data analysis software, HPLC−DAD raw data are often only used by expert users to check for peak purity or to identify a compound by comparison of the UV−Vis spectrum with a reference spectrum. Most workflows access chromatogram analysis results by vendor software in the form of peak tables. In extreme cases, a full HPLC−DAD raw data array is recorded only to extract one value out of a peak table, e.g., the area of the product signal, while all of the other information is discarded. This is incongruous with modern data-centered automation and digitalization approaches.

Commercial software solutions from Virscidian (Analytical Studio[26]) or ACD/Labs (Katalyst D2D, Spectrus[27]) have already filled the gap of modern multivariate raw data analysis. However, as commercial products, they provide limited flexibility in workflow implementation. For example, Virscidian had to implement a construct called *expressions* in their software to allow the user to extract relevant information in a customizable and flexible manner. The analytical chemistry community is also adopting multivariate data analysis, but code availability is limited.[28−31] For example, Arase et al. explored with the Shimadzu Corporation as an HPLC instrument vendor the potential of HPLC−DAD data in the context of peak deconvolution.[32]

In this work, we present the open-source Python project MOCCA (Multivariate Online Contextual Chromatographic Analysis), which enables the direct processing and analysis of HPLC−DAD raw data in Python, the de facto standard programming/scripting language for data science projects in chemistry.[33−36] As a ready-to-use Python package, it is easily implemented into existing automated and nonautomated workflows. By making the Pythonic library of data analysis toolkits accessible, MOCCA enables its users to develop new and powerful data analysis features. Here, we present a peak deconvolution feature that allows for automated deconvolution and quantification of known signals which overlap with signals of unexpected impurities. This overcomes a common limitation of available commercial software: the requirement for manual control of the automatic integration routine to account for overlapping peaks. With this feature implemented, MOCCA could play a major role toward autonomous laboratories by providing open actionable analytics, i.e., enabling data-based decisions without human intervention or control by putting HPLC−DAD data in the correct context for analysis.[19]

Other open-source toolkits exist for chromatographic data analysis, e.g., HappyTools[37] and Aston[38] in Python or chromatographR[39] in R. The authors of the Alsace package for R emphasized the potential of DAD data for metabolomics profiling.[40] Notably, Jason Hein and co-workers recently developed a Python-based automated data processing routine and made the code available online.[41] However, all these efforts do not make consistent use of the multidimensionality of the HPLC−DAD data or are developed with a specific use case in mind so that they are not ready-to-use for a synthetic chemist. We envision MOCCA to serve as a basis for a joint community effort toward an open multivariate analytical raw data analysis toolkit.

MOCCA serves as a plug-and-play module and is not restrained by a specific project scope. To highlight MOCCA's general applicability and versatility for chemical reaction analysis, we introduce and investigate MOCCA's data analysis features in four different case studies. First, the features are validated using a large set of simulated chromatograms including overlapping signals for a quantitative investigation of the peak deconvolution feature. Then, the potential of MOCCA and its peak deconvolution feature is highlighted in an experimental reaction kinetics study on a Knoevenagel condensation reaction. In the third study, MOCCA is employed in a closed-loop process optimization for the alkylation of 2-pyridone where the peak deconvolution feature keeps the optimization cycle running despite the signal of an unexpected side product overlaps with the product signal. Finally, a newly developed cyanation of aryl halides is presented and categorical reaction parameters are screened on a well plate with MOCCA tracking all known and unknown signals.

2. METHODS

Our proposed analytical workflow employing the MOCCA package in automated, semiautomated, or nonautomated workflows is shown in Figure 1. In research laboratories, HPLC systems from a number of different vendors are used in combination with corresponding vendor-specific control software. The HPLC−DAD raw data (time−wavelength absorbance array) are typically stored in proprietary formats inaccessible to the user. To obtain open, nonproprietary HPLC−DAD raw data files, each of the softwares has its own native raw data export routine. Therefore, MOCCA includes raw data parsers for data exported from Agilent's ChemStation, Shimadzu's LabSolutions, and Water's Empower software. However, we highly encourage, if possible, exporting to standardized and metadata-enriched data formats such as the Allotrope data format,[42] for which a parser is implemented in the MOCCA package. These standardized data formats ensure

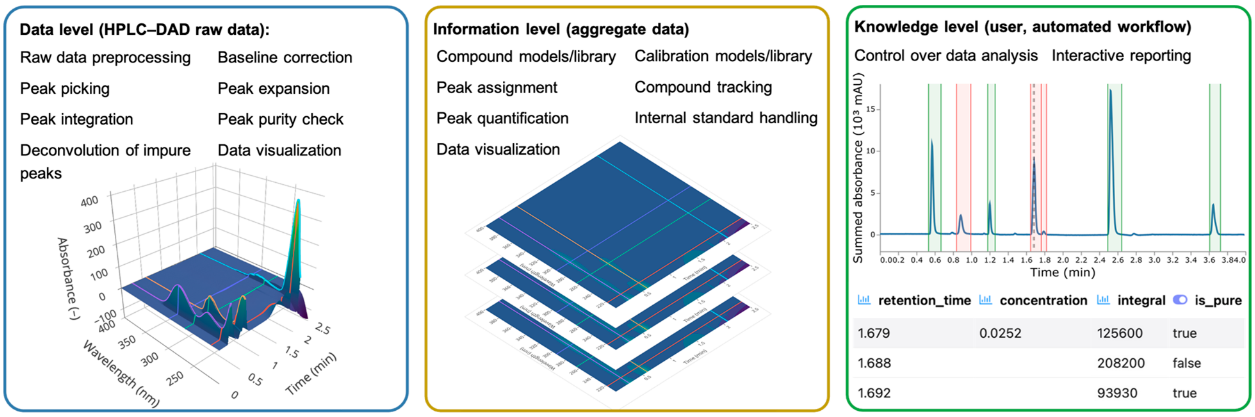


Figure 2. Summary of the data analysis features implemented in MOCCA.

the implementation of FAIR (findability, accessibility, interoperability, reuse) principles in analytical data and promote reuse of data for future scientific projects (details in SI section S2).[43]

After parsing the exported data in Python, HPLC−DAD raw data sets are analyzed by MOCCA. An automated procedure extracts relevant information for a specific scientific question from the information-rich analysis results. The obtained structured (tabular) data sets are used for data-based decision making.

A summary of the single data analysis features of the MOCCA package is presented in Figure 2 (details in SI sections S3 and S4). The features are assigned to three hierarchy levels: the raw data level, the aggregate data level, and the user or automated workflow interaction level. On the raw data level, most features are known from common vendor software and include raw data preprocessing with baseline correction as well as peak picking and integration. Other features like the algorithms for automated peak purity checking and peak deconvolution can complement vendor software capabilities. These two features are discussed and validated in detail in the following sections. On the aggregate data level, information is created by analyzing data sets in context to each other. By mimicking and automating routine steps a scientist would perform in the lab, compound and calibration libraries are created to allow for peak assignment and peak quantification. Moreover, MOCCA allows for the automated handling of internal standards for retention time correction as well as for relative quantification. Finally, interaction with the tool takes place on the highest hierarchy level, which provides control over certain settings of the data analysis and provides interactive reports on the analysis. The reports include the most crucial information for the user, such as chromatogram visualizations and peak tables (examples in chromatogram report in HTML SI files).

3. RESULTS AND DISCUSSION

3.1. Validation of Data Analysis Features in Simulated Chromatograms.

Collecting large-scale experimental HPLC−DAD data for validation is time-consuming and inefficient. Moreover, experimental data sets with overlapping signals do not provide a ground truth against which deconvolution results can be quantitatively compared. To solve this problem, we turned to the Chromatography Analysis and Design Toolkit (CADET), a tool that simulates retention processes on LC separation columns.[44] With CADET, a wide variety of elution profiles can be simulated (including nongaussian shapes) while taking into account nonlinear retention effects of coeluting analytes.

Table 1. Results of the MOCCA Analysis of Synthetic HPLC−DAD Data Assigned to the Following Categories: (i) Retention Profiles of Main Compound and Impurity Were Baseline-Separated and Analyzed Correctly; (ii) Overlapping Retention Profiles Where the Peak Deconvolution Feature Was Triggered and the Main Compound Was Identified and Quantified; (iii) Peak Deconvolution Feature Was Triggered for the Overlapping Signal but the Main Compound Could Not Be Identified; (iv) the Signals Were Overlapping but Were Not Labelled As Impure by the Peak Purity Checker

	UV−Vis spectral similarity		
Result category	High	Medium	Low
(i)	86	86	86
(ii)	794	868	890
(iii)	2	0	0
(iv)	118	46	24

To imitate a real situation and systematically explore the limitations of MOCCA's peak deconvolution feature, chromatograms with two compounds (a known main compound and an unknown impurity) were generated in-silico. The resulting retention profiles were enriched with compound-specific UV−Vis spectra to obtain synthetic HPLC−DAD data sets. The similarity of UV−Vis spectra of the main compound and the impurity were varied in three levels of correlation coefficients r, high ($r \approx 0.86$), medium ($r \approx 0.47$), and low ($r \approx -0.06$). To obtain statistically relevant results, we simulated 1000 different chromatograms and enriched them with UV−Vis spectra of each similarity level resulting in 3000 HPLC−DAD data sets (details in SI section S7). These synthetic raw data were fed into MOCCA for data analysis. This allowed for the testing and validation of all other features shown in Figure 2. We provide the simulated data sets in the Supporting Information as a benchmark for future developments.

The obtained results were assigned to four possible categories: (i) separate peaks where the simulated pair of retention profiles is baseline-separated and the peak purity checker labels the two peaks as pure; (ii) successful deconvolution of an overlapping signal where the main compound is correctly assigned and quantified; (iii)

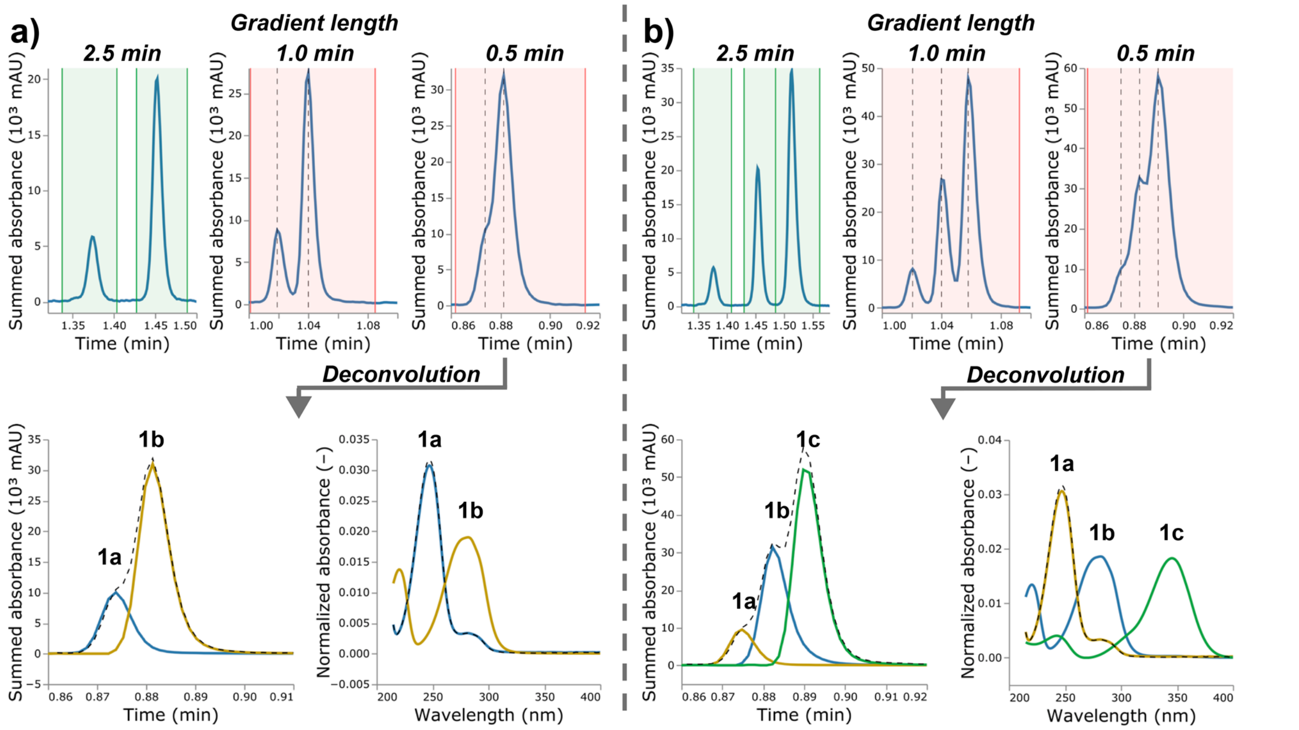


Figure 3. (a) Results of the competition experiment with two benzaldehydes (**1a** and **1b**). (b) Results of the competition experiment with three benzaldehydes (**1a**−**c**). *Top*: Chromatographic signals of the benzaldehydes using different gradient lengths. MOCCA indicates results of purity checks (green passed, red failed) and centers of retention profiles modeled by the deconvolution algorithm (vertical black dashed lines). *Bottom*: Deconvolution results of the overlapping signal recorded with a gradient length of 0.5 min. The modeled retention profiles (left, colored lines) described the retention profile of the impure peak (black dashed line). The modeled UV−Vis traces (right, colored lines) correspond to the UV−Vis spectra of the benzaldehydes as exemplified for **1a** (black dashed line).

unsuccessful deconvolution where the peak is labeled as impure but the deconvolution feature is not able to assign any deconvoluted component to the main compound; (iv) no trigger of the peak deconvolution feature due to the peak purity check returning a false positive result. In general, cases (i) and (ii) are considered as desired outcomes, while cases (iii) and (iv) are considered as misinterpretations (examples in SI section S7). Table 1 summarizes the obtained results highlighting that a vast majority of the simulated cases were processed correctly while almost all of the failing cases are attributed to category (iv).

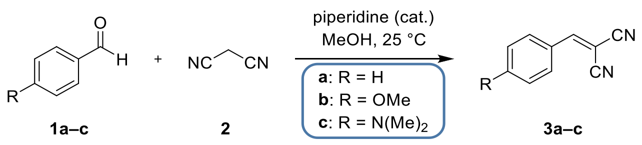
Cases of the category (iv) are not attributed to a failure of the peak deconvolution feature, but rather to a permissive peak purity checker returning false positive outcomes on strongly coeluting signals. These cases cannot be solved analytically, and the only solution would be the development of an HPLC method with higher chromatographic resolution to separate (at least partially) the elution profiles. MOCCA enables a shift toward shorter gradient times and faster sample processing, but the category (iv) failure rate shows that the user is still required to have expertise in HPLC method development to balance method time vs chromatographic resolution.[45]

For a quantitative investigation of the devonvolution results, we looked at the results assigned to category (ii) and compared them to the ground truth. For all three levels of spectral similarity, the median quantification error was smaller than 2%, while the third quartile error ranged around 6% (examples and details in SI section S7). The results obtained validate that MOCCA's deconvolution feature works robustly enough for typical lab screenings, but should be treated with caution for regulated environments and process development scenarios where lower margins of error are required.

3.2. Kinetics Study of Knoevenagel Condensation Reactions.

A reaction kinetics study of a Knoevenagel condensation, a well-established test reaction,[46,47] was conducted to highlight the potential of MOCCA's peak deconvolution feature. Benzaldehydes (**1a**−**c**) were simultaneously reacted with malononitrile (**2**) to their corresponding benzylidenemalonitriles (**3a**−**c**) in the same reaction mixture (Scheme 1).

Scheme 1. Knoevenagel Condensation Reactions of Benzaldehyde (1a), 4-Methoxybenzaldehyde (1b) and 4-(Dimethylamino)benzaldehyde (1c) with Malononitrile (2) in Methanol (MeOH) to Yield Benzylidenemalonitriles 3a−c



Reactions were performed in an HPLC vial in a temperature-controlled (25 °C) autosampler, and reaction progress was followed via reversed-phase HPLC with different gradient lengths. Five different HPLC methods were developed with gradient lengths of 0.5, 0.75, 1.0, 1.5, and 2.5 min (water/acetonitrile 95:5 → 0:100 v/v) to induce different degrees of overlap between the substrate signals. For quantification, calibration curves were recorded for all substrates with all

HPLC methods. These measurements were used to validate the quantification features of MOCCA in the case of pure and baseline-separated signals against traditional manual data analysis. The results of both analysis methods correlated very precisely (details in SI section S5).

Two competition experiments were performed: malononitrile (**2**) was reacted with two (**1a**, **1b**), and with three (**1a**–**c**) benzaldehyde substrates, respectively. For data analysis, benzaldehyde (**1a**) was treated as the main compound, i.e., only its calibration runs were added to MOCCA for quantitative analysis while the functionalized benzaldehydes **1b** and **1c** were treated as "unknown" impurities. Figure 3a and b illustrate results from the two competition experiments. The top panels show the different degrees of signal overlap induced by the gradient variation. Here, the peak purity check feature correctly labeled the peaks as pure for the long gradient (green background area) and correctly labeled the overlapping peaks as impure for the short gradients (red background area). In the latter cases, the peak deconvolution feature was triggered. As a first step of the deconvolution routine, a principal component analysis is performed on the absorbance array of an impure peak to estimate the number of overlapping components. With this number as an input, a newly developed iterative algorithm using parallel factor analysis (PARAFAC)[48] is employed for deconvolution (details in SI section S4). The bottom panels in Figure 3 show the deconvolution results for the peaks recorded with a gradient length of 0.5 min.

To investigate the ability of MOCCA to automatically recognize impure peaks and decompose a known signal from coeluting impurities, reaction progress was followed by sampling out of the same reaction vessel repeating each of the five HPLC methods iteratively (details in SI section S6). The resulting reaction kinetics plots of the main compound benzaldehyde (**1a**) are shown in Figure 4 and exhibit the expected second-order kinetics.[46,49] As expected from the simulation study, the results of chromatograms with baseline-separated signals agree with the results of chromatograms where signals were heavily overlapping. The deconvolution feature successfully identified the benzaldehyde (**1a**) signal in all given impure peaks and returned modeled peaks for quantification.

3.3. Closed-Loop Optimization of the Alkylation of 2-Pyridone.

Closed-loop optimization studies have gained tremendous attention in recent years due to their relevance for chemical discovery as well as process optimization.[50−54] In such closed-loop processes, the optimization platform runs without human intervention and control of HPLC data analysis. Here, the peak purity check and peak deconvolution feature of MOCCA are of particular interest. Overlapping peaks or inaccurate integration routines (examples in SI section S5) lead to wrong analytical results fed back to the experimental design algorithm. The setup of the closed-loop optimization platform of this study is shown in Figure 5. For the experimental design, we employed a Python package called Experimental Design via Bayesian Optimization (EDBO) published by Doyle and co-workers.[55] The suggested optimization parameter values were fed into a LabVIEW program controlling a microfluidic droplet platform, which was developed in the Jensen group for the simultaneous screening of both categorical and continuous reaction parameters.[56−60] After reaction completion in an oscillatory droplet reactor, the droplet was diluted with acetonitrile and moved to an internal injection valve (0.02 μL injection volume) to inject a sample directly on a reversed-phase separation column of the HPLC system. The HPLC system automatically exported HPLC–DAD raw data for MOCCA data analysis after each run. The optimization objective, as well as process control parameters were extracted from the MOCCA analysis results via a project-specific script.

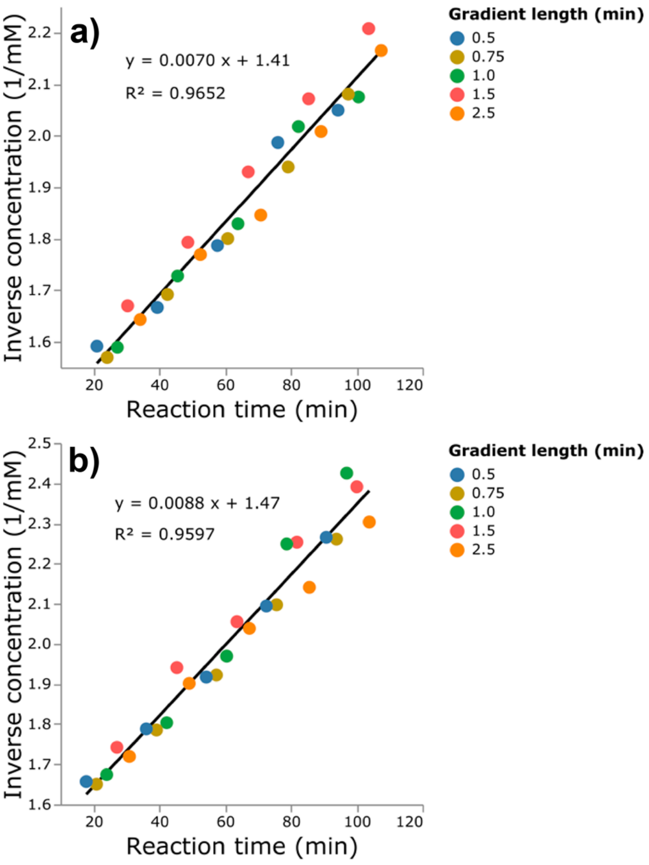


Figure 4. Second-order reaction kinetics plots of benzaldehyde (**1a**) in the Knoevenagel condensation recorded with five different HPLC methods employing varying gradient lengths. (a) Competition experiment with two benzaldehydes (**1a** and **1b**). (b) Competition experiment with with three benzaldehydes (**1a**–**c**).

As a test reaction, we examined the alkylation of 2-pyridone (**4**) using 1-iodobutane (**5**) to yield the two regioisomers 1-butylpyridone (**6**) and 2-butoxypyridine (**7**). As shown in Scheme 2, the optimization was performed on two continuous variables, the reaction time (10–60 min) and temperature (35–100 °C). Additionally, two categorical optimization parameters were screened: the base (1,8-diazabicyclo[5.4.0]-undec-7-ene (DBU), 1,1,3,3-tetramethylguanidine (TMG), *N*,*N*-diisopropylethylamine (DIPEA)), and the solvent (*n*-butanol, *N*,*N*-dimethylformamide (DMF), toluene). The maximization of the yield of 1-butylpyridone (**6**) served as the objective function for the optimization. For quantification, the desired product **6** was calibrated relative to an internal standard in an automatic fashion by the platform (details in SI section S8).

The optimization cycle was run with a batch size of one, i.e., the feedback loop was closed after each experiment. At any point during the optimization campaign, the user was able to extract MOCCA reports to follow the optimization process. Figure 6a summarizes the results of the optimization campaign. The optimal conditions found for the reaction were DBU in

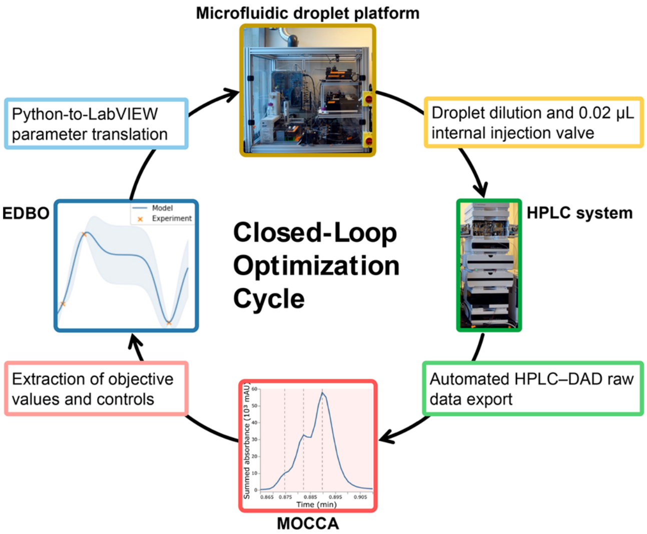
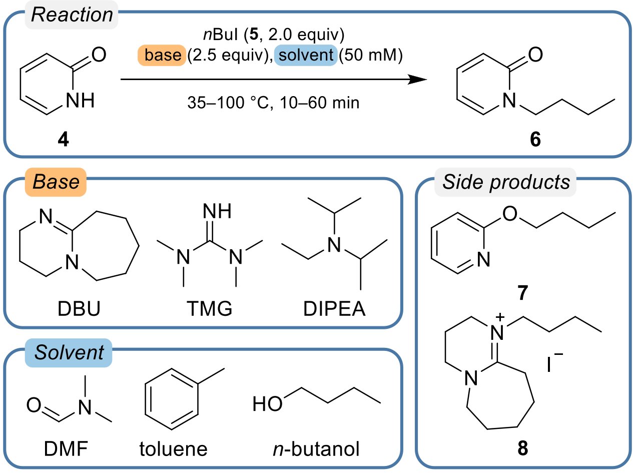

Microfluidic droplet platform
Python-to-LabVIEW parameter translation
Droplet dilution and 0.02 µL internal injection valve
EDBO
Model
Experiment
Closed-Loop Optimization Cycle
HPLC system
Extraction of objective values and controls
Automated HPLC–DAD raw data export
Summed absorbance (10³ mAU)
Time (min)
MOCCA

Figure 5. Closed-loop optimization cycle employed in this work. *Blue*: Experimental Design via Bayesian Optimization (EDBO) Python package from the Doyle group[55] and translation of the suggested parameters to a LabVIEW experimental protocol. *Yellow*: Experimental execution by a microfluidic reactor platform employing an oscillatory droplet reactor design. 0.02 μL HPLC samples are taken out of the droplet after diltution with acetonitrile. *Green*: HPLC system with a photodiode array detector (DAD) and an automated HPLC−DAD raw data export routine. *Red*: Data analysis by the MOCCA tool and a project-specific script for the extraction of objective values and process control values.

Scheme 2. Optimization Campaign on the Alkylation of 2-Pyridone (4) with 1-Iodobutane (5) Yielding 1-Butylpyridone (6)[a]



[a]The domain space of the optimization campaign spans over two continuous variables, reaction time (low boundary: 10 min, high boundary: 60 min), and temperature (low boundary: 35 °C, high boundary: 100 °C), as well as two categorical variables, identity of base (DBU, TMG, DIPEA) and solvent (DMF, toluene, *n*-butanol). The objective value of the optimization is the yield of **6**. Two side products were identified with 2-butoxypyridine (**7**) and butylated DBU (**8**).

toluene for 60 min at 100 °C. We validated the obtained optimization results with batch reactions that screened all categorical parameter combinations at 35 and 100 °C (details in SI section S8). For all reactions with DBU, the HPLC signal of an unexpected side product, butylated DBU (**8**), started overlapping with the signal of the calibrated product **6**, whose yield served as the objective value for the optimization. Figure 6b shows the chromatogram of the reaction at optimal conditions with the impure peak at ∼1.7 min resulting from an overlap of signals from **6** and **8**. As shown in Figure 6c, MOCCA was able to deconvolute this impure peak in an automated fashion and to feed back corrected yields to the design algorithm EDBO. This highlights MOCCA's ability to keep closed-loop cycles running even when unexpected coelution of calibrated signals occurs in the HPLC analysis.

3.4. Palladium-Catalyzed Cyanation of Aryl Halides.

With this study, we highlight MOCCA's application for the analysis of HPLC−DAD data originating from a novel palladium-catalyzed cyanation of aryl halides, where side products were unknown. Palladium-catalyzed cyanation reactions are a prominent and well-investigated reaction class.[61−63] They proceed via oxidative addition of an aryl halide to a Pd(0)/ligand complex, subsequent halide/cyanide exchange, followed by a reductive elimination which closes the catalytic cycle.[64] A particular challenge with this reaction class resides in the rapid deactivation of the catalytically active palladium species in the presence of excess amounts of cyanide.[65,66] To overcome this issue, many procedures were developed with the aim of keeping a low effective concentration of cyanide in solution. Common strategies include the use of hardly soluble metal salts,[67−69] employing cyanide transfer agents,[70] and the slow addition of trimethylsilyl cyanide[71] or acetone cyanohydrin.[72,73] The use of butyronitrile in combination with a nickel catalyst allows for cyanide release through a reverse hydrocyanation reaction.[74]

Giumond et al.[73] developed a protocol for palladium and nickel catalyzed cyanation reactions to overcome upscaling issues associated with the use of metal cyanides under heterogeneous conditions.[68,75−77] A homogeneous reaction is obtained by adding acetone cyanohydrin via syringe pump to a solution of the substrate, a palladium catalyst, a ligand, and *N*,*N*-diisopropylethylamine (DIPEA) in isopropyl alcohol or *n*-butanol (Scheme 3a).[73] Based on these results, we envisioned to make use of *O*-protected cyanohydrins as cyanation reagents which release cyanide in situ upon deprotection (Scheme 3b). This approach maintains a fully homogeneous liquid system but the need for slow reagent addition is avoided.

To investigate our proposed synthetic strategy, we prepared a number of protected cyanide-releasing agents **10a**−**10g** (Figure 7b). In situ deprotection by transesterification or TMS cleavage yields acetone cyanohydrin or lactonitrile which rapidly eliminate the cyanide required for cross-coupling. We screened suitable reaction conditions for the conversion of 2-chlorotoluene (**9**) to *o*-tolunitrile (**11**) in a 96 well plate (Figure 7a) by combining these reagents with one of three different ligands (Figure 7d), XPhos, *t*BuXPhos, or CM-Phos, and one of four different bases (Figure 7c), DBU, TMG, 4-(dimethylamino)pyridine (DMAP), or DIPEA (details in SI section S9). The choice of ligands and $[Pd(cinnamyl)Cl]_2$ as the catalyst precursor was based on previous literature reports.[73,78−80]

After the reaction was run under the given conditions and subsequent internal standard addition, samples of the reactions were subjected to HPLC analysis. The HPLC−DAD raw data were exported as text files and parsed for subsequent MOCCA analysis, which took ∼10 min on a standard personal laptop for the whole data set including 42 deconvolutions of impure

https://doi.org/10.1021/acscentsci.2c01042

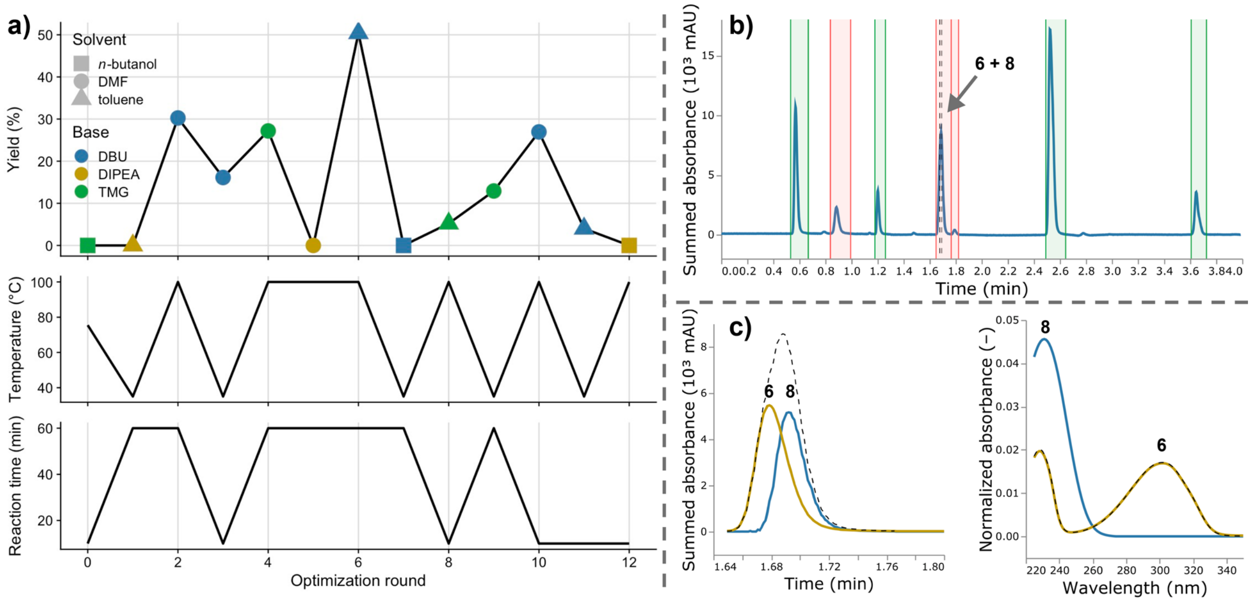
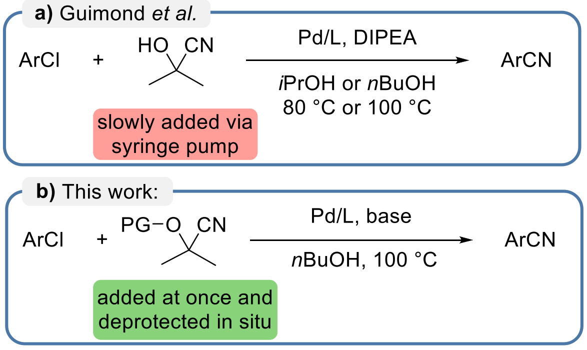


Figure 6. Results of the closed-loop optimization on the alkylation of 2-pyridone (**4**). (a) Objective values as a function of optimizer choices in each round. *Top*: Objective value (yield of **6**) with marker shape indicating the chosen solvent and marker color indicating the chosen base; *middle*: Chosen reaction temperature; *bottom*: Chosen reaction time. (b) Chromatogram of the reaction under optimal conditions with an impure product peak (~1.7 min). (c) Modeled retention profiles (dashed line: impure peak) and UV−Vis spectra (dashed line: reference UV−Vis spectrum of (**6**) of the product **6** (yellow) and the unexpected impurity **8** (blue).

Scheme 3. (a) Palladium-Catalyzed Cyanation of Aryl Chlorides Developed by Guimond et al. Based on the Slow Addition of Acetone Cyanohydrin via Syringe Pump.[73] (b) Newly Developed Cyanation Method Using Protected Cyanohydrins (PG: protecting group) for in Situ Release of Cyanide



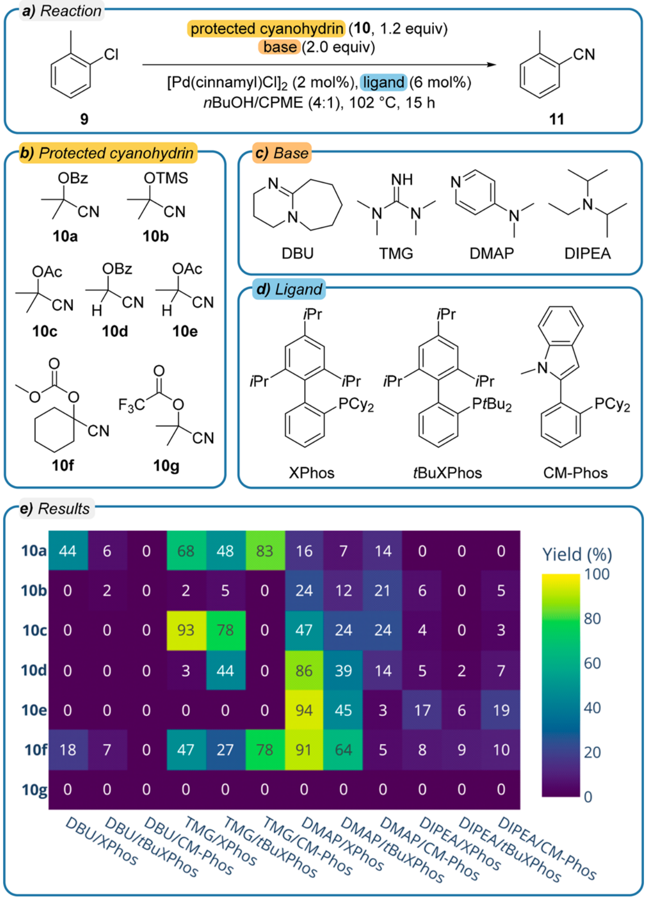
peaks. The MOCCA analysis results enabled following product and substrate concentrations and, importantly, unknown signals over the data sets, thus supporting the identification of side products and impurities (example in SI section S9). These data were used for heatmap visualization in Python using standard toolkits (Plotly[81]). For example, the obtained yields of *o*-tolunitrile (**11**) are visualized based on their location in the well plate (Figure 7e). This again highlights the potential of moving HPLC−DAD data analysis to Python with its powerful package library for data analysis and visualization.

As discussed above, a successful reaction requires the release of cyanide anions to proceed at a rate that is sufficient to be productive but not outpace the catalyst turnover. This rate is controlled through the rates of deprotection of the cyanide-releasing agents **10**, which were examined experimentally for a better understanding of our results (details in SI section S9). The screening provided three parameter combinations with yields >90% indicating a good harmonization between cyanide release and turnover, the TMG/XPhos base−ligand combination with the precursor **10c**, DMAP/XPhos with **10e**, and DMAP/XPhos with **10f**. The outcome of these experiments together with a selection of the other experiments were verified by repeating the reactions in standard reaction flasks (details in SI section S9). For other parameter combinations, e.g., when using trifluoroacetylated cyanohydrin **10g**, the release of cyanide is too fast, leading to a quick catalyst deactivation. In contrast, a slow release of cyanide is observed with the use of DIPEA, a weak base, leading to low conversions (detailed mechanistic discussion in SI section S9).

HPLC analysis represents a typical bottleneck in well plate-based screenings. Typically, HPLC methods are developed to be as short as possible for maximum throughput while resolving all known compounds. When screening categorical variables like ligands or bases, unexpected side products often overlap with known signals in the chromatogram. This also happened in the described screening campaign, but MOCCA reliably deconvoluted these overlapping peaks and enabled an efficient data analysis without the need for HPLC method optimization or resorting to multiplexing techniques (examples and details in SI section S9).[82,83]

4. CONCLUSIONS

In this work, we have presented MOCCA, an open-source Python project, for the comprehensive analysis of HPLC−DAD raw data. Compared to typical data analysis methodologies on one signal wavelength, the analysis of the full time−wavelength absorbance array gives multiple advantages. These include robust peak assignment and quantification, as well as peak purity checks and the deconvolution of overlapping peaks. We investigated MOCCA in four case studies, (i) a simulation study, (ii) a reaction kinetics study, (iii) a closed-loop optimization, (iv) a well plate screening and demon-



	DBU/XPhos	DBU/tBuXPhos	DBU/CM-Phos	TMG/XPhos	TMG/tBuXPhos	TMG/CM-Phos	DMAP/XPhos	DMAP/tBuXPhos	DMAP/CM-Phos	DIPEA/XPhos	DIPEA/tBuXPhos	DIPEA/CM-Phos
10a	44	6	0	68	48	83	16	7	14	0	0	0
10b	0	2	0	2	5	0	24	12	21	6	0	5
10c	0	0	0	93	78	0	47	24	24	4	0	3
10d	0	0	0	3	44	0	86	39	14	5	2	7
10e	0	0	0	0	0	0	94	45	3	17	6	19
10f	18	7	0	47	27	78	91	64	5	8	9	10
10g	0	0	0	0	0	0	0	0	0	0	0	0

Figure 7. (a) Reaction conditions for the well plate screening of the cyanation of 2-chlorotoluene (**9**) yielding *o*-tolunitrile (**11**) using palladium(π-cinnamyl) chloride dimer as the catalyst precursor. (b) Screened *O*-protected cyanohydrins. (c) Screened bases. (d) Screened ligands. (e) Yield of *o*-tolunitrile (**11**) in dependency on the employed protected cyanide-releasing agent **10** as well as the chosen ligand and base. CPME: cyclopentyl methyl ether; Bz: benzoyl group; Ac: acetyl group; TMS: trimethylsilyl group.

strated MOCCA's broad applicability and the benefit of moving chromatographic data analysis to an open environment like Python.

In this spirit, we envision MOCCA becoming a community project with a significant user base eager to adapt, curate, and further advance the tool. With community support, MOCCA can overcome limitations of vendor software especially with regard to FAIR data principles and implementation in automated workflows. The development of additional data analysis features such as the implementation of a mass spectrometry module could extend the scope of the tool by adding orthogonal analysis dimensions. Another interesting development could be a connection MOCCA to chemical structure representations, or even to chemical reaction entries in electronic lab notebooks. This would make synthetic chemistry data and the corresponding analytical data directly accessible for machine learning in data science applications.

To enable new users to implement MOCCA easily in their laboratories, we packaged MOCCA and published it in the Python Package Index (PyPI). For a quick start, example JupyterLab notebooks together with the corresponding HPLC−DAD data sets are provided in the notebooks folder of the package's GitHub repository.[84] This includes a tutorial as well as the complete data analysis of the well plate screening presented in this manuscript.

■ ASSOCIATED CONTENT

Data Availability Statement

The full Python base code is available through GitHub at https://github.com/HaasCP/mocca. The MOCCA package is available on the PyPI server (https://pypi.org/project/mocca/) and MOCCA can be installed following the documentation at https://mocca.readthedocs.io/en/latest/readme.html. Simulated HPLC−DAD data sets saved as MOCCA campaigns, which were used for validation and benchmarking, are available at 10.5281/zenodo.7406829.

SI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.2c01042.

Additional details to all presented case studies, description how to extract HPLC−DAD raw data from vendor control software of major vendors, technical details to MOCCA's data analysis features, NMR spectra of *O*-protected cyanohydrins (PDF)

MOCCA reports for the data analysis of the well plate screening (cyanation of aryl halides) (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Julius Hillenbrand − *Chemical & Pharmaceutical Development, Bayer AG, Pharmaceuticals Division, 42117 Wuppertal, Germany*; orcid.org/0000-0002-2646-1302; Email: julius.hillenbrand@bayer.com

Giulio Volpin − *Research and Development, Small Molecules Technologies, Bayer AG, Crop Science Division, 65926 Frankfurt am Main, Germany*; Email: giulio.volpin@bayer.com

Klavs F. Jensen − *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*; orcid.org/0000-0001-7192-580X; Email: kfjensen@mit.edu

Authors

Christian P. Haas − *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Research and Development, Small Molecules Technologies, Bayer AG, Crop Science Division, 65926 Frankfurt am Main, Germany*; orcid.org/0000-0002-9457-8019

Maximilian Lübbesmeyer − *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Research and Development, Small Molecules Technologies, Bayer AG, Crop Science Division, 65926 Frankfurt am Main, Germany*

Edward H. Jin − *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*; orcid.org/0000-0001-6011-5211

Matthew A. McDonald − *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*; orcid.org/0000-0002-9444-3253

Brent A. Koscher − *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge,* *Massachusetts 02139, United States;* orcid.org/0000-0001-8233-0852

Nicolas Guimond − *Research and Development, Small Molecules Technologies, Bayer AG, Crop Science Division, 40789 Monheim am Rhein, Germany;* orcid.org/0000-0001-5258-2557

Laura Di Rocco − *Chemical & Pharmaceutical Development, Bayer AG, Pharmaceuticals Division, 13353 Berlin, Germany*

Henning Kayser − *Research and Development, Small Molecules Technologies, Bayer AG, Crop Science Division, 40789 Monheim am Rhein, Germany*

Samuel Leweke − *Applied Mathematics, Bayer AG, Enabling Functions Division, 51368 Leverkusen, Germany;* orcid.org/0000-0001-9471-4511

Sebastian Niedenführ − *Research and Development, Computational Life Science, Bayer AG, Crop Science Division, 40789 Monheim am Rhein, Germany*

Rachel Nicholls − *Research and Development, Computational Life Science, Bayer AG, Crop Science Division, 40789 Monheim am Rhein, Germany*

Emily Greeves − *Research and Development, Small Molecules Technologies, Bayer AG, Crop Science Division, 65926 Frankfurt am Main, Germany*

David M. Barber − *Research and Development, Weed Control Chemistry, Bayer AG, Crop Science Division, 65926 Frankfurt am Main, Germany;* orcid.org/0000-0001-9906-1695

Complete contact information is available at:
https://pubs.acs.org/10.1021/acscentsci.2c01042

Author Contributions

Author contribution roles using CRediT.[85] Conceptualization: CPH, ML, NG, GV, KFJ; data curation: CPH, EHJ; formal analysis: CPH, ML, EHJ; funding acquisition: NG, GV, KFJ; investigation: CPH, ML, EHJ, MAM, BK, EG; methodology: CPH, ML, EHJ, JH, GV, KFJ; project administration: CPH, ML, NG, DMB, JH, GV, KFJ; resources: DMB, NG, JH, GV, KFJ; software: CPH, EHJ (MOCCA), LDR, HK (allotrope parser), SL (CADET), SN, RN, ML (implementation of EDBO); supervision: CPH, JH, GV, KFJ; validation: CPH, ML, EHJ, RN, SN; visualization: CPH, ML, EHJ; writing−original draft: CPH, ML; writing−review and editing: CPH, ML, EHJ, MAM, BAK, NG, DMB, JH, GV, KFJ.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to acknowledge Romaric Gérardy, Anirudh Nambiar, Andrew Zahrt, Florian Häse, Matthias Westphal, Moll Glass, and Thorsten Gressling for helpful discussions as well as Natalie Eyke, Brooke Jin, Travis Hart, Prajwal Mahesh, Jan-Erik Seelbach, Andre Althoff, Michael Krause, ADIRO Automatisierungstechnik GmbH (Ute Ferlein, Jürgen Helmich, Volker Hödl), and Angi GmbH (Arno Dorschner) for technical assistance. This work was supported by Life Science Collaboration Program (Project MIC DROP) of Bayer AG and partially by the DARPA Accelerated Molecular Discovery programs under contract HR00111920025 (MAM and BAK).

■ REFERENCES

(1) Andersson, S.; Armstrong, A.; Björe, A.; Bowker, S.; Chapman, S.; Davies, R.; Donald, C.; Egner, B.; Elebring, T.; Holmqvist, S.; Inghardt, T.; Johannesson, P.; Johansson, M.; Johnstone, C.; Kemmitt, P.; Kihlberg, J.; Korsgren, P.; Lemurell, M.; Moore, J.; Pettersson, J. A.; Pointon, H.; Pontén, F.; Schofield, P.; Selmi, N.; Whittamore, P. Making Medicinal Chemistry More Effective-Application of Lean Sigma to Improve Processes, Speed and Quality. *Drug Discovery Today* **2009**, *14* (11−12), 598−604.

(2) Whitesides, G. M. Reinventing Chemistry. *Angew. Chemie - Int. Ed.* **2015**, *54* (11), 3196−3209.

(3) Matlin, S. A.; Mehta, G.; Hopf, H.; Krief, A. One-World Chemistry and Systems Thinking. *Nat. Chem.* **2016**, *8* (5), 393−398.

(4) Marion, P.; Bernela, B.; Piccirilli, A.; Estrine, B.; Patouillard, N.; Guilbot, J.; Jérôme, F. Sustainable Chemistry: How to Produce Better and More from Less? *Green Chem.* **2017**, *19* (21), 4973−4989.

(5) Keijer, T.; Bakker, V.; Slootweg, J. C. Circular Chemistry to Enable a Circular Economy. *Nat. Chem.* **2019**, *11* (3), 190−195.

(6) Kulik, H. J.; Sigman, M. S. Advancing Discovery in Chemistry with Artificial Intelligence: From Reaction Outcomes to New Materials and Catalysts. *Acc. Chem. Res.* **2021**, *54* (10), 2335−2336.

(7) Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part I: Progress. *Angew. Chemie - Int. Ed.* **2020**, *59*, 22858.

(8) Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part II: Outlook. *Angew. Chemie - Int. Ed.* **2020**, *59*, 23414.

(9) Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; dos Passos Gomes, G.; Zepel, T.; Sigman, M. S.; Aspuru-Guzik, A.; Hein, J. E. Data-Science Driven Autonomous Process Optimization. *Commun. Chem.* **2021**, *4* (1), 1−12.

(10) Coley, C. W.; Thomas, D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; John Hart, A.; Jamison, T. F.; Jensen, K. F. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* **2019**, *365* (6453), eaax1566.

(11) Trobe, M.; Burke, M. D. The Molecular Industrial Revolution: Automated Synthesis of Small Molecules. *Angew. Chemie - Int. Ed.* **2018**, *57* (16), 4192−4214.

(12) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7* (10), 1622−1637.

(13) Schneider, P.; Walters, W. P.; Plowright, A. T.; Sieroka, N.; Listgarten, J.; Goodnow, R. A.; Fisher, J.; Jansen, J. M.; Duca, J. S.; Rush, T. S.; Zentgraf, M.; Hill, J. E.; Krutoholow, E.; Kohler, M.; Blaney, J.; Funatsu, K.; Luebkemann, C.; Schneider, G. Rethinking Drug Design in the Artificial Intelligence Era. *Nat. Rev. Drug Discovery* **2020**, *19* (5), 353−364.

(14) dos Passos Gomes, G.; Pollice, R.; Aspuru-Guzik, A. Navigating through the Maze of Homogeneous Catalyst Design with Machine Learning. *Trends Chem.* **2021**, *3* (2), 96−110.

(15) Struble, T. J.; Alvarez, J. C.; Brown, S. P.; Chytil, M.; Cisar, J.; Desjarlais, R. L.; Engkvist, O.; Frank, S. A.; Greve, D. R.; Griffin, D. J.; Hou, X.; Johannes, J. W.; Kreatsoulas, C.; Lahue, B.; Mathea, M.; Mogk, G.; Nicolaou, C. A.; Palmer, A. D.; Price, D. J.; Robinson, R. I.; Salentin, S.; Xing, L.; Jaakkola, T.; Green, W. H.; Barzilay, R.; Coley, C. W.; Jensen, K. F. Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *J. Med. Chem.* **2020**, *63* (16), 8667−8682.

(16) Christensen, M.; Yunker, L. P. E.; Shiri, P.; Zepel, T.; Prieto, P. L.; Grunert, S.; Bork, F.; Hein, J. E. Automation Isn't Automatic. *Chem. Sci.* **2021**, *12* (47), 15473−15490.

(17) Molga, K.; Szymkuć, S.; Gołębiowska, P.; Popik, O.; Dittwald, P.; Moskal, M.; Roszak, R.; Mlynarski, J.; Grzybowski, B. A. A Computer Algorithm to Discover Iterative Sequences of Organic Reactions. *Nat. Synth.* **2022**, *1* (1), 49−58.

(18) Gromski, P. S.; Granda, J. M.; Cronin, L. Universal Chemical Synthesis and Discovery with 'The Chemputer.'. *Trends Chem.* **2020**, *2* (1), 4−12.

(19) Shi, Y.; Prieto, P. L.; Zepel, T.; Grunert, S.; Hein, J. E. Automated Experimentation Powers Data Science in Chemistry. *Acc. Chem. Res.* **2021**, *54* (3), 546−555.

(20) Fitzner, M.; Wuitschik, G.; Koller, R. J.; Adam, J. M.; Schindler, T.; Reymond, J. L. What Can Reaction Databases Teach Us about Buchwald-Hartwig Cross-Couplings? *Chem. Sci.* **2020**, *11* (48), 13085−13093.

(21) Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143* (45), 18820−18826.

(22) Campos, K. R.; Coleman, P. J.; Alvarez, J. C.; Dreher, S. D.; Garbaccio, R. M.; Terrett, N. K.; Tillyer, R. D.; Truppo, M. D.; Parmee, E. R. The Importance of Synthetic Chemistry in the Pharmaceutical Industry. *Science* **2019**, *363* (6424), eaat0805.

(23) Aldeghi, M.; Häse, F.; Hickman, R. J.; Tamblyn, I.; Aspuru-Guzik, A. Golem: An Algorithm for Robust Experiment and Process Optimization. *Chem. Sci.* **2021**, *12* (44), 14792−14807.

(24) Beker, W.; Roszak, R.; Wolos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki-Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144* (11), 4819−4827.

(25) Tortorella, S.; Cinti, S. How Can Chemometrics Support the Development of Point of Need Devices? *Anal. Chem.* **2021**, *93* (5), 2713−2722.

(26) *Virscidian Analytical Studio, accessed June 2022*. https://www.virscidian.com.

(27) *ACD/Labs, accessed June 2022*. https://www.acdlabs.com.

(28) Patel, D. C.; Wahab, M. F.; O'Haver, T. C.; Armstrong, D. W. Separations at the Speed of Sensors. *Anal. Chem.* **2018**, *90* (5), 3349−3356.

(29) Molenaar, S. R. A.; Dahlseid, T. A.; Leme, G. M.; Stoll, D. R.; Schoenmakers, P. J.; Pirok, B. W. J. Peak-Tracking Algorithm for Use in Comprehensive Two-Dimensional Liquid Chromatography − Application to Monoclonal-Antibody Peptides. *J. Chromatogr. A* **2021**, *1639*, 461922.

(30) Niezen, L. E.; Schoenmakers, P. J.; Pirok, B. W. J. Critical Comparison of Background Correction Algorithms Used in Chromatography. *Anal. Chim. Acta* **2022**, *1201*, 339605.

(31) Schmidt, B.; Jaroszewski, J. W.; Bro, R.; Witt, M.; Stærk, D. Combining PARAFAC Analysis of HPLC-PDA Profiles and Structural Characterization Using HPLC-PDA-SPE-NMR-MS Experiments: Commercial Preparations of St. John's Wort. *Anal. Chem.* **2008**, *80* (6), 1978−1987.

(32) Arase, S.; Horie, K.; Kato, T.; Noda, A.; Mito, Y.; Takahashi, M.; Yanagisawa, T. Intelligent Peak Deconvolution through In-Depth Study of the Data Matrix from Liquid Chromatography Coupled with a Photodiode Array Detector Applied to Pharmaceutical Analysis. *J. Chromatogr. A* **2016**, *1469*, 35−47.

(33) O'Boyle, N. M.; Morley, C.; Hutchison, G. R. Pybel: A Python Wrapper for the OpenBabel Cheminformatics Toolkit. *Chem. Cent. J.* **2008**, *2* (1), 1−7.

(34) Gressling, T. *Data Science in Chemistry*; Walter de Gruyter GmbH: Berlin/Boston, 2021.

(35) Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: An Orchestration Software to Democratize Autonomous Discovery. *PLoS One* **2020**, *15* (4), e0229862.

(36) Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. Self-Referencing Embedded Strings (SELFIES): A 100% Robust Molecular String Representation. *Mach. Learn. Sci. Technol.* **2020**, *1* (4), 045024.

(37) Jansen, B. C.; Hafkenscheid, L.; Bondt, A.; Gardner, R. A.; Hendel, J. L.; Wuhrer, M.; Spencer, D. I. R. HappyTools: A Software for High-Throughput HPLC Data Processing and Quantitation. *PLoS One* **2018**, *13* (7), e0200280.

(38) Bovee, R. *Aston*, accessed June 2022.https://github.com/bovee/Aston

(39) Bass, E. *chromatographR: chromatographic data analysis toolset*, accessed June 2022. https://cran.r-project.org/package=chromatographR/.

(40) Wehrens, R.; Carvalho, E.; Fraser, P. D. Metabolite Profiling in LC−DAD Using Multivariate Curve Resolution: The Alsace Package for R. *Metabolomics* **2015**, *11* (1), 143−154.

(41) Liu, J.; Sato, Y.; Yang, F.; Kukor, A. J.; Hein, J. E. An Adaptive Auto-Synthesizer Using Online PAT Feedback to Flexibly Perform a Multistep Reaction. *Chemistry−Methods* **2022**, *2*, 1−9.

(42) Jones, P.-J.; Roberts, J. M.; Vanderwall, D. E.; Vergis, J. M. Unlocking the Power of Data. *LCGC North Am.* **2015**, *33*, 270−281.

(43) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J. W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; t Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S. A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; Van Der Lei, J.; Van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. Comment: The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 1−9.

(44) Leweke, S.; von Lieres, E. Chromatography Analysis and Design Toolkit (CADET). *Comput. Chem. Eng.* **2018**, *113*, 274−294.

(45) Mattrey, F. T.; Makarov, A. A.; Regalado, E. L.; Bernardoni, F.; Figus, M.; Hicks, M. B.; Zheng, J.; Wang, L.; Schafer, W.; Antonucci, V.; Hamilton, S. E.; Zawatzky, K.; Welch, C. J. Current Challenges and Future Prospects in Chromatographic Method Development for Pharmaceutical Research. *TrAC - Trends Anal. Chem.* **2017**, *95*, 36−46.

(46) Hruby, S. L.; Shanks, B. H. Acid-Base Cooperativity in Condensation Reactions with Functionalized Mesoporous Silica Catalysts. *J. Catal.* **2009**, *263* (1), 181−188.

(47) Haas, C. P.; Müllner, T.; Kohns, R.; Enke, D.; Tallarek, U. High-Performance Monoliths in Heterogeneous Catalysis with Single-Phase Liquid Flow. *React. Chem. Eng.* **2017**, *2* (4), 498−511.

(48) Escandar, G. M.; Olivieri, A. C. Multi-Way Chromatographic Calibration—A Review. *J. Chromatogr. A* **2019**, *1587*, 2−13.

(49) Haas, C. P.; Tallarek, U. Kinetics Studies on a Multicomponent Knoevenagel−Michael Domino Reaction by an Automated Flow Reactor. *ChemistryOpen* **2019**, *8* (5), 606−614.

(50) Holmes, N.; Akien, G. R.; Blacker, A. J.; Woodward, R. L.; Meadows, R. E.; Bourne, R. A. Self-Optimisation of the Final Stage in the Synthesis of EGFR Kinase Inhibitor AZD9291 Using an Automated Flow Reactor. *React. Chem. Eng.* **2016**, *1* (4), 366−371.

(51) Clayton, A. D.; Schweidtmann, A. M.; Clemens, G.; Manson, J. A.; Taylor, C. J.; Niño, C. G.; Chamberlain, T. W.; Kapur, N.; Blacker, A. J.; Lapkin, A. A.; Bourne, R. A. Automated Self-Optimisation of Multi-Step Reaction and Separation Processes Using Machine Learning. *Chem. Eng. J.* **2020**, *384*, 123340.

(52) Porwol, L.; Kowalski, D. J.; Henson, A.; Long, D. L.; Bell, N. L.; Cronin, L. An Autonomous Chemical Robot Discovers the Rules of Inorganic Coordination Chemistry without Prior Knowledge. *Angew. Chemie - Int. Ed.* **2020**, *59* (28), 11256−11261.

(53) Breen, C. P.; Nambiar, A. M. K.; Jamison, T. F.; Jensen, K. F. Ready, Set, Flow! Automated Continuous Synthesis and Optimization. *Trends Chem.* **2021**, *3* (5), 373−386.

(54) Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Becker, J. S.; Aspuru-Guzik, A. Closed-Loop Discovery Platform Integration Is Needed for Artificial Intelligence to Make an Impact in Drug Discovery. *Expert Opin. Drug Discovery* **2019**, *14* (1), 1−4.

(55) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89−96.

(56) Reizman, B. J.; Jensen, K. F. Simultaneous Solvent Screening and Reaction Optimization in Microliter Slugs. *Chem. Commun.* **2015**, *51* (68), 13290−13293.

(57) Reizman, B. J.; Wang, Y. M.; Buchwald, S. L.; Jensen, K. F. Suzuki-Miyaura Cross-Coupling Optimization Enabled by Automated Feedback. *React. Chem. Eng.* **2016**, *1* (6), 658−666.

(58) Hwang, Y. J.; Coley, C. W.; Abolhasani, M.; Marzinzik, A. L.; Koch, G.; Spanka, C.; Lehmann, H.; Jensen, K. F. A Segmented Flow Platform for On-Demand Medicinal Chemistry and Compound Synthesis in Oscillating Droplets. *Chem. Commun.* **2017**, *53* (49), 6649−6652.

(59) Coley, C. W.; Abolhasani, M.; Lin, H.; Jensen, K. F. Material-Efficient Microfluidic Platform for Exploratory Studies of Visible-Light Photoredox Catalysis. *Angew. Chemie - Int. Ed.* **2017**, *56* (33), 9847−9850.

(60) Baumgartner, L. M.; Coley, C. W.; Reizman, B. J.; Gao, K. W.; Jensen, K. F. Optimum Catalyst Selection over Continuous and Discrete Process Variables with a Single Droplet Microfluidic Reaction Platform. *React. Chem. Eng.* **2018**, *3* (3), 301−311.

(61) Sundermeier, M.; Zapf, A.; Beller, M. Palladium-Catalyzed Cyanation of Aryl Halides: Recent Developments and Perspectives. *Eur. J. Inorg. Chem.* **2003**, *2003* (19), 3513−3526.

(62) Anbarasan, P.; Schareina, T.; Beller, M. Recent Developments and Perspectives in Palladium-Catalyzed Cyanation of Aryl Halides: Synthesis of Benzonitriles. *Chem. Soc. Rev.* **2011**, *40* (10), 5049−5067.

(63) Neetha, M.; Afsina, C. M. A.; Aneeja, T.; Anilkumar, G. Recent Advances and Prospects in the Palladium-Catalyzed Cyanation of Aryl Halides. *RSC Adv.* **2020**, *10* (56), 33683−33699.

(64) Takagi, K.; Okamoto, T.; Sakakibara, Y.; Ohno, A.; Oka, S.; Hayama, N. Nucleophilic Displacement Catalyzed by Transition Metal. I. General Consideration of the Cyanation of Aryl Halides Catalyzed by Palladium(II). *Bull. Chem. Soc. Jpn.* **1975**, *48* (11), 3298−3301.

(65) Dobbs, K. D.; Marshall, W. J.; Grushin, V. V. Why Excess Cyanide Can Be Detrimental to Pd-Catalyzed Cyanation of Haloarenes. Facile Formation and Characterization of [Pd(CN)3-(H)] 2 and [Pd(CN)3(Ph)]2-. *J. Am. Chem. Soc.* **2007**, *129* (1), 30−31.

(66) Erhardt, S.; Grushin, V. V.; Kilpatrick, A. H.; Macgregor, S. A.; Marshall, W. J.; Roe, D. C. Mechanisms of Catalyst Poisoning in Palladium-Catalyzed Cyanation of Haloarenes. Remarkably Facile C-N Bond Activation in the [(Ph3P) 4Pd]/[Bu4N]+ CN- System. *J. Am. Chem. Soc.* **2008**, *130* (14), 4828−4845.

(67) Takagi, K.; Okamoto, T.; Sakakibara, Y.; Ohno, A.; Oka, S.; Hayama, N. Nucleophilic Displacement Catalyzed by Transition Metal. III. Kinetic Investigation of the Cyanation of Iodobenzene Catalyzed by Palladium(II). *Bull. Chem. Soc. Jpn.* **1976**, *49* (11), 3177−3180.

(68) Ushkov, A. V.; Grushin, V. V. Rational Catalysis Design on the Basis of Mechanistic Understanding: Highly Efficient Pd-Catalyzed Cyanation of Aryl Bromides with NaCN in Recyclable Solvents. *J. Am. Chem. Soc.* **2011**, *133* (28), 10999−11005.

(69) Tschaen, D. M.; Desmond, R.; King, A. O.; Fortin, M. C.; Pipik, B.; King, S.; Verhoeven, T. R. An Improved Procedure for Aromatic Cyanation. *Synth. Commun.* **1994**, *24* (6), 887−890.

(70) Schareina, T.; Zapf, A.; Beller, M. Potassium Hexacyanoferrate-(II)—a New Cyanating Agent for the Palladium-Catalyzed Cyanation of Aryl Halides. *Chem. Commun.* **2004**, *4* (12), 1388−1389.

(71) Sundermeier, M.; Mutyala, S.; Zapf, A.; Spannenberg, A.; Beller, M. A Convenient and Efficient Procedure for the Palladium-Catalyzed Cyanation of Aryl Halides Using Trimethylsilylcyanide. *J. Organomet. Chem.* **2003**, *684* (1−2), 50−55.

(72) Sundermeier, M.; Zapf, A.; Beller, M. A Convenient Procedure for the Palladium-Catalyzed Cyanation of Aryl Halides. *Angew. Chemie Int. Ed.* **2003**, *42* (14), 1661−1664.

(73) Burg, F.; Egger, J.; Deutsch, J.; Guimond, N. A Homogeneous Method for the Conveniently Scalable Palladium- and Nickel-Catalyzed Cyanation of Aryl Halides. *Org. Process Res. Dev.* **2016**, *20* (8), 1540−1545.

(74) Yu, P.; Morandi, B. Nickel-Catalyzed Cyanation of Aryl Chlorides and Triflates Using Butyronitrile: Merging Retro-Hydrocyanation with Cross-Coupling. *Angew. Chemie Int. Ed.* **2017**, *56* (49), 15693−15697.

(75) Marcantonio, K. M.; Frey, L. F.; Liu, Y.; Chen, Y.; Strine, J.; Phenix, B.; Wallace, D. J.; Chen, C. Y. An Investigation into Causes and Effects of High Cyanide Levels in the Palladium-Catalyzed Cyanation Reaction. *Org. Lett.* **2004**, *6* (21), 3723−3725.

(76) Weissman, S. A.; Zewge, D.; Chen, C. Ligand-Free Palladium-Catalyzed Cyanation of Aryl Halides. *J. Org. Chem.* **2005**, *70* (4), 1508−1510.

(77) Magano, J.; Dunetz, J. R. Large-Scale Applications of Transition Metal-Catalyzed Couplings for the Synthesis of Pharmaceuticals. *Chem. Rev.* **2011**, *111* (3), 2177−2250.

(78) Senecal, T. D.; Shu, W.; Buchwald, S. L. A General, Practical Palladium-Catalyzed Cyanation of (Hetero)Aryl Chlorides and Bromides. *Angew. Chemie Int. Ed.* **2013**, *52* (38), 10035−10039.

(79) Cohen, D. T.; Buchwald, S. L. Mild Palladium-Catalyzed Cyanation of (Hetero)Aryl Halides and Triflates in Aqueous Media. *Org. Lett.* **2015**, *17* (2), 202−205.

(80) Yeung, P. Y.; So, C. M.; Lau, C. P.; Kwong, F. Y. A Mild and Efficient Palladium-Catalyzed Cyanation of Aryl Chlorides with K4[Fe(CN)6]. *Org. Lett.* **2011**, *13* (4), 648−651.

(81) Plotly Technologies Inc. *Collaborative data science*, accessed August 2022. https://plot.ly.

(82) Welch, C. J.; Gong, X.; Schafer, W.; Pratt, E. C.; Brkovic, T.; Pirzada, Z.; Cuff, J. F.; Kosjek, B. MISER Chromatography (Multiple Injections in a Single Experimental Run): The Chromatogram Is the Graph. *Tetrahedron Asymmetry* **2010**, *21* (13−14), 1674−1681.

(83) Santanilla, A. B.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.; Schneeweis, J.; Berritt, S.; Shi, Z.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D. Nanomole-Scale High-Throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347* (6217), 443−448.

(84) Haas, C. P. *MOCCA GitHub repository*, 2022. https://github.com/HaasCP/mocca.

(85) *Contributor Roles Taxonomy*, accessed August 2022. https://credit.niso.org.